QH573.C38
mi
453108

STIC-ILL

From: Wessendorf,Teresa
Sent: Saturday, June 28, 2003 2:13 PM
To: STIC-ILL
Subject: 09/990091

Please forward:
1.Cell, 38, 309-316 (1984).
2. J. Cell.Evol. 108, 229-241 (1989).
Txs.
T. Wessendorf
Art Unit 1639
Rm. 2B17
MailRm. 3B01
308-3967

Cell, Vol. 38, 309–316, August 1984, Copyright 1984 by MIT

0092-8674/84/080309-08 $02.00/0

Trypan some mRNAs Share a Common 5′ Spliced L ad r Sequence

Marilyn Parsons, Richard G. Nelson, Kenneth P. Watkins, and Nina Agabian
University of Washington
Department of Biochemistry SJ-70
Seattle, Washington 98195

Summary

A 5′-terminal leader sequence of 35 nucleotides was found to be present on multiple trypanosome RNAs. Based on its representation in cDNA libraries, we estimate that many, if not all, trypanosome mRNAs contain this leader. This same leader was originally identified on mRNAs encoding the molecules responsible for antigenic variation, variant surface glycoproteins. Studies of selected cDNAs containing this leader sequence revealed that leader-containing transcripts can be stage-specific, stage-regulated, or constitutive. They can be abundant or rare, and transcribed from single or multigene families. No linkage between the genomic leader sequences and the structural gene exons was observed. Possible mechanisms by which the leader sequences are added to trypanosome mRNAs are discussed.

Introduction

The surface of pathogenic African trypanosomes is covered with a densely packed coat composed of a single protein species, the variant surface glycoprotein (VSG). The VSG gene repertoire of the trypanosome contains some 300 to 1000 genes (Van der Ploeg et al., 1982a), yet only one VSG gene at a time is expressed. Switching transcription from one VSG gene to another leads to antigenic variation, allowing the parasite to evade the immune response of its mammalian host. The result is the relapsing parasitemia of African sleeping sickness. Some of the molecular and biological aspects of antigenic variation have been recently reviewed (Parsons et al., 1984a).

The molecular mechanism that assures the transcriptional activation of a single VSG gene remains unknown. Comparisons of the genomic contexts of VSG genes have revealed that certain VSG genes undergo duplication when activated (Hoeijmakers et al., 1980a; Pays et al., 1981a; Longacre et al., 1983; Parsons et al., 1983). The new expression-linked copy is located at a different site in the genome and is transcribed (Pays et al., 1981b; Bernards et al., 1981). Other VSG genes appear to be activated in situ, as they undergo no detectable alteration in genomic organization when expressed (Young et al., 1982). Transcriptionally active VSG genes invariably reside in regions of the genome relatively devoid of restriction enzyme recognition sites (the "barren" regions) (Van der Ploeg et al., 1982b) and close to what appears to be a telomere (DeLange and Borst, 1982; Williams et al., 1982). These data suggest that some aspects of the genomic location of VSG genes may be important for their activation. However, analyses of restriction enzyme sites located 5′ to the barren regions indicates that there are several sites in the genome from which VSG genes can be transcribed (Longacre et al., 1983; Myler et al., submitted; Allison et al., submitted). Furthermore, while a VSG gene must reside in one of these sites in order to be transcribed, the opposite is not true; occupation of a particular site does not guarantee transcription (Buck et al., 1984; Allison et al., submitted).

Whether transcribed from duplication- or nonduplication-activated genes or from genes residing in similar or distinct genomic locations, all VSG mRNAs share the same 35 nucleotide sequence at their 5′ terminus (Van der Ploeg et al., 1982c; Boothroyd and Cross, 1982). This untranslated sequence is not encoded by DNA contiguous to the structural gene, and therefore has been termed the spliced leader (SL). Sequences encoding the SL are repeated 100–200 times in the genome of Trypanosoma brucei, and each resides in a 1.4 kb unit monomer (DeLange et al., 1983; Nelson et al., 1983). The vast majority of these units are directly and tandemly repeated to form a large array(s). However, a few 1.4 kb units, with their resident SL sequences, are dispersed from the tandem array (Nelson et al., 1983; Parsons et al., 1984b) and are termed orphons according to the nomenclature of Childs et al. (1981). Surprisingly, neither the large array nor the orphons are detectably linked (i.e., within 50 kb) to active VSG genes. Nevertheless, the tandem array of SL reiteration units has been proposed to mark the 5′ boundary of the VSG expression site (DeLange et al., 1983). Immediately 5′ to the SL in the 1.4 kb repeat unit are sequences resembling eucaryotic RNA polymerase II promoters (DeLange et al., 1983); according to the multiple promoter hypothesis above, these function in the initiation of VSG gene transcription. Only that VSG gene placed downstream from the array through DNA rearrangement would be transcribed, thus only a single VSG gene at a time would be expressed.

Certain clues, however, indicate that the function of the SL repeat is more complex. First, the SL is transcribed not only by the mammalian bloodstream stage of the parasite, which expresses VSG and undergoes antigenic variation, but also by procyclic culture forms (analogous to the insect midgut stage of the trypanosome life cycle), which do not express VSG (DeLange et al., 1983; Parsons et al., 1984b). Second, sequences homologous to the SL are highly reiterated in the genomes of certain trypanosomatids that do not undergo antigenic variation, such as Trypanosoma cruzi, the intracellular parasite that causes Chagas disease, and Leptomonas collosoma, a parasite of insects (Nelson et al., 1984). These findings suggest that although the SL is used by African trypanosomes for VSG expression, it is an ancient sequence that may fulfill other more fundamental functions in gene expression in these organisms. This would predict the presence of the SL on other trypanosome transcripts as well as on VSG mRNA. T test this hypothesis we have screened cDNA libraries made from T. brucei bloodstream or procyclic RNA for clones hybrid-

izing to a synthetic probe complementary to 22 nucleotides of the 35 nucleotide SL. Recombinant clones which did not encode VSG, but which contained the SL, were detected in both libraries. The expression of the corresponding genes is regulated differently from VSG genes; they are transcribed during both bloodstream and procyclic stages. As with VSG genes however, the SL is not encoded contiguously to these various structural genes, and is apparently derived from a transcript originating elsewhere in the genome. Our data suggest that rather than providing for transcriptional exclusivity in VSG gene expression, the SL plays a universal role in trypanosome gene expression.

Results

Isolation of cDNA Clones Containing the SL

Only bloodstream-stage trypanosomes synthesize VSG mRNA (Agabian et al., 1980; Overath et al., 1983; Parsons et al., 1984b), and then, of the thousand-odd VSG genes, each variant antigen type (VAT) transcribes only the VSG gene that is ultimately expressed as its surface coat (Hoeijmakers et al., 1980b; Pays et al., 1980). Thus, in Figure 1A, when total RNA is fractionated by agarose gel electrophoresis, transferred to nitrocellulose membranes, and hybridized to a ^{32}P-labeled probe (Northern analysis), a prominent species is detected by a cDNA encoding VSG 3 in RNA isolated from bloodstream-stage cells of VAT 3 (lane B). This RNA, of about 1.9 kb, and a putative precursor of about 4 kb (seen only upon overexposure of the autoradiogram), are not observed in procyclic RNA (lane P) nor in bloodstream-stage RNA isolated from other VATs (not shown). As previously reported (Parsons et al., 1984b), when the same RNAs are hybridized to a probe complementary to 22 nucleotides of the 35 nucleotide SL (see Figure 2) a very different pattern is obtained (Figure 1B). Here, a smear of RNAs ranging in size from 0.5 to over 6 kb is detected in both procyclic (lane P) and bloodstream stage (lane B). In particular, a 1.9 kb species is seen in VAT 3 RNA (lane B). Ribosomal RNAs (their position indicated by the hash marks) do not hybridize to the synthetic probe. The SL-containing RNAs are not simply transcripts of the genomic 1.4 kb repeat unit in which the SL resides, as rehybridization of the same blot to a clone containing the 1.4 kb genomic repeat reveals no hybridizing RNA (not shown). Since many of the RNAs revealed by the SL probe were smaller than the mature VSG mRNA, it seemed unlikely that they represented VSG mRNA processing intermediates. We therefore hypothesized that the smear of RNAs hybridizing to the synthetic probe represents transcripts derived from other genes that also use the SL or an SL-like sequence.

As a first step in testing this hypothesis, we used the SL probe to screen two cDNA libraries, one made from RNA isolated from VAT 5 and one made from RNA isolated from procyclic cells derived from VAT 5. In the procyclic library of 1000 clones, 24 clones were found that hybridized with the synthetic probe. Of approximately 600 recombinant clones in the VAT 5 library, 25 VSG-encoding cDNAs were detected, but only one of these contained 5′-terminal sequences as defined by hybridization to the SL probe. Seventeen other clones also hybridized with the SL probe. By analogy with VSG transcripts, the SL should be located at the 5′ terminus of transcripts giving rise to these clones. If only 5% of our cDNAs contain 5′-terminal sequences (as is the case for the VSG clones), then the numbers cited above indicate that approximately half of the clones in the library are derived from transcripts containing the SL. Thus, as predicted from the Northern analyses, SL-containing transcripts are abundant in stages of the life cycle when VSG is (bloodstream) or is not (procyclic) expressed. Most of the SL-containing cDNAs were represented only once in the libraries, implying that the complexity of such transcripts is high.

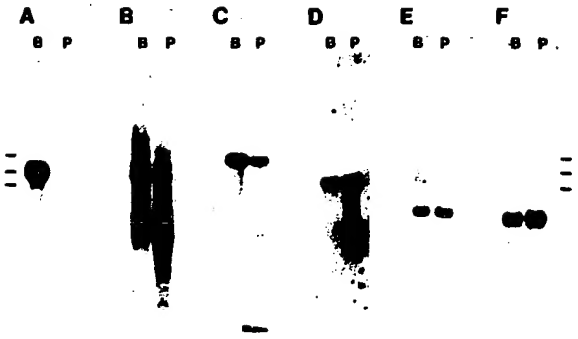

Figure 1. Northern Analyses

Total bloodstream VAT 3 RNA (B) and procyclic RNA (P) were fractionated on adjacent lanes (in quadruplicate) of a single gel. Four blots were prepared and hybridized to the probes listed below. When blots were re-used, we removed the previous probe with the appropriate denaturation conditions, verified by autoradiography. The probes used were as follows: (A) VSG 3 cDNA; (B) 22-mer SL; (C) pSLc1 cDNA; (D) pSLc2 cDNA; (E) pSLc4 cDNA; (F) pSLC3 cDNA.

Nucleotide Sequence Analysis of SL cDNAs

To determine whether the hybridizing sequences contained in these cDNAs were identical with the SL found on VSG mRNAs (Figure 2) or were divergent SL-like sequences, and to search for possible protein-coding sequences, the hybridizing regions of four cDNAs (two from the bloodstream-stage library, and two from the procyclic library) were cloned into M13 bacteriophage and their nucleotide sequences were determined. As shown in Figure 2, each clone contained a complete or partial SL. In the case of pSLc1 and pSLc3, the SL was truncated; the cDNA clone contained 23 and 26 bp, respectively, of the 35 nucleotide SL sequence. We interpret these to be incomplete copies of their corresponding transcripts. Since the synthetic probe is complementary to the 3′ portion of the SL, and since the cDNA and probe sequences match perfectly, these clones were easily detected by the SL probe. Clone pSLc4 contained a complete SL and, 5′ to that, 3 nucleotides that correspond to those immediately 5′ to the SL in the genomic repeat unit. These data suggest that sequences 5′ to the SL may be transcribed, as has

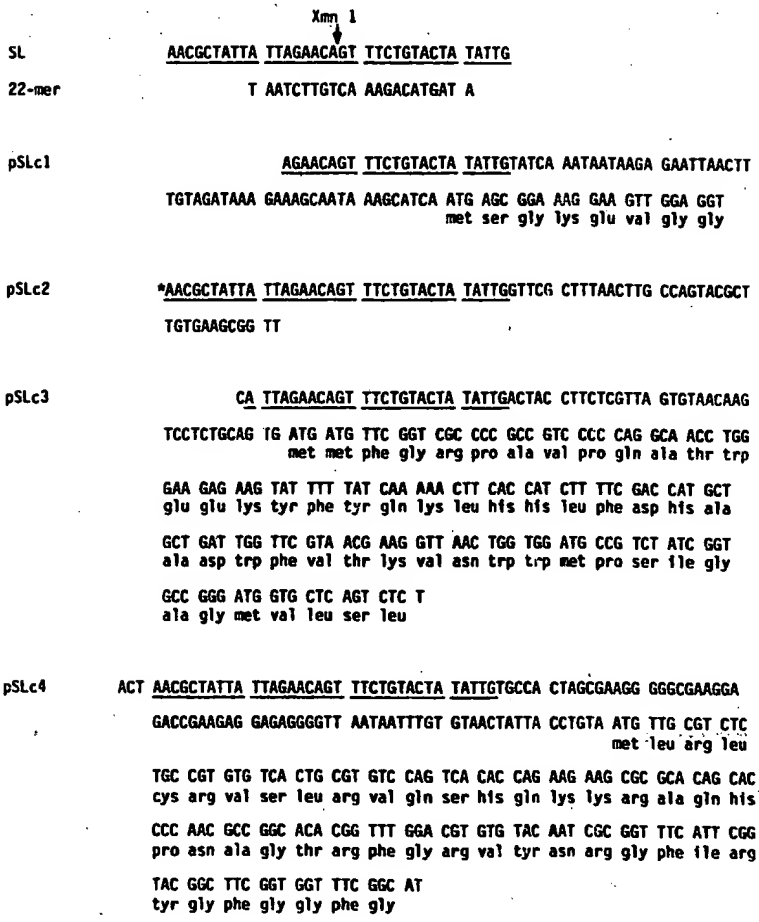

Figure 2. Nucleotide Sequence Analysis

The sequence of the 35 nucleotide SL and the sequence of the 22 nucleotide probe complementary to the SL are shown at the top. Below are shown partial sequences of several cDNAs that hybridize to the SL probe (the SL portion is underlined). Putative reading frames are translated into amino acid sequences. The asterisk (*) indicates that another 92 bp of sequence was found 5′ to the SL (see text).

been proposed by Boothroyd and Cross (1982). The fourth clone, pSLc2, contained a complete and perfect SL. An additional 93 bp (indicated by an asterisk in Figure 2) are found 5′ to the SL in this clone. Subsequent hybridization analysis suggested that this 93 bp sequence is derived from sequences within the structural gene, implying that its appearance 5′ to the SL resulted from an artifact of cDNA cloning. Michiels et al. (1983) have described a VSG cDNA containing sequences 5′ to the SL.

The start codons of VSG mRNAs are located some 30 to 80 nucleotides downstream from the SL. We further analyzed these clones for the presence of a start codon and the existence of a possible reading frame. Our results indicate that, like the VSG leaders, these leaders are composite, containing the 5′ SL followed by additional noncoding sequences. For example, the first ATG triplet is encountered 53 bp downstream from the SL in pSLc1. In pSLc3, a start codon is found 37 bp downstream from the SL, followed by a reading frame extending at least 159 nucleotides. pSLc4 contains a reading fram that starts 72 nucleotides downstream from the SL, extending at least 135 nucleotides. This region encodes a 45 amino acid polypeptide rich in basic amino acids. Thus there is no reason to suspect that these cDNAs are not derived from bona fide mRNAs since they contain protein-coding reading frames.

Expression of Genes that Use the SL

The one well-studied class of genes that employs the SL, VSG genes, shows stage-specific expression. It was therefore interesting to determine whether stage-specific expression was a general feature of genes that use the SL. The transcripts that correspond to several of the SL-containing cDNAs were examined by Northern analysis of RNA isolated from the two easily obtained stages of the trypanosome life cycle: bloodstream and procyclic stages. The RNA blots shown in Figures 1A and 1B were rehybridized with SL cDNA probes corresponding to single-copy genes (see below). In each case one major RNA species was detected, ranging in size from 1.3 to 2.3 kb, depending on the cDNA used as probe (Figure 1, C–F). The two clones from the procyclic library, pSLc3 (Figure 1F) and pSLc4 (Figure 1E), also detect additional higher molecular weight RNAs in both bloodstream and procyclic samples. For example, pSLc4 hybridizes primarily to a 1.4 kb RNA, but also detects 1.75 and 2.0 kb RNAs. (These are more clearly visible on longer autoradiographic exposures.) Unlike VSG genes, all of these genes are tran-

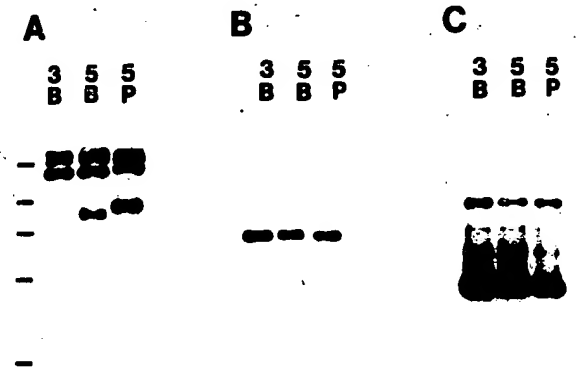

Figure 3. Genomic Southerns

DNA isolated from trypanosomes of VAT 3 (3B), VAT 5 (5B), or procyclic cells derived from VAT 5 (5P) was cleaved with Eco RV. Two micrograms of each sample was electrophoresed in triplicate on an agarose gel and transferred to nitrocellulose paper and hybridized with nick-translated SL cDNAs. The hash marks on the left indicate the position of DNA markers at 23 kb, 9.6 kb, 6.6 kb, 4.3 kb, and 2.3 kb. Probes were: A, VSG 5 DNA; B, pSLc2; C, pSLc5.

scribed during both bloodstream and procyclic phases of the trypanosome life cycle (compare lanes P and B). One of the cDNAs isolated from the bloodstream-stage library, pSLc1, detects a transcript that appears more abundant in the bloodstream than in procyclic stage (Figure 1C), while pSLc3, isolated from the procyclic library, detects a transcript that is more abundant in procyclic RNA (Figure 1F). Since the SL cDNA probes had specific activities similar to that of the VSG cDNA probe, but required autoradiographic exposures some five to ten times longer, it seems that the abundance of each of these transcripts is quite low compared to that of VSG gene transcripts (Figure 1A).

Genomic Organization of Genes That Use the SL

Genomic Southern analysis using SL cDNA probes reveals several classes of genes whose transcripts contain the SL. These experiments were performed under conditions too stringent for stable hybridization of the 35 nucleotide SL with its corresponding genomic sequences. Figure 3 shows Eco RV-cleaved genomic DNA from bloodstream-stage VATs 3 (3B) and 5 (5B) and from the procyclic population obtained by in vitro differentiation of VAT 5 (5P). When hybridized with a VSG 5 cDNA probe, the hallmark properties of many VSG genes are visible (Figure 3A): the presence of a multigene family, an expression-linked copy (the extra gene copy in VAT 5 as compared to VAT 3), and the retention of the expression-linked copy by procyclic cells (Parsons et al., 1984b). In addition, the expression-linked copy, as well as one other member of the VSG 5 gene family, resides on restriction fragments that show variations in size. Such variation is characteristic of VSG genes that are located adjacent to a telomere and is correlated with cell growth or division (Bernards et al., 1983).

In contrast, in this and many other restriction digests, pSLc2 hybridizes to a single genomic sequence (Figure 3B), indicating that pSLc2 detects a single-copy gene. These data also suggest that there are no large introns within the portion of the gene detected by the cDNA probe (an intron between the SL and the structural gene would not be detected, since the SL sequence does not hybridize to its corresponding genomic sequences under these conditions). Unlike VSG genes, this gene shows no variation between VATs or life-cycle stages. Of six SL-containing cDNA clones used as probes in Southern analyses, four appeared to detect single-copy, nonvarying genes. The remaining two cDNAs detected nonvarying multigene families such as that shown in Figure 3C. Since all trypanosome telomeric sequences thus far studied show variation in length over time, it appears that, unlike many VSG genes, these other genes that employ the SL do not reside adjacent to a telomere. Again, unlike active VSG genes, the other genes that employ the SL are not situated in DNA devoid of restriction sites (see also Figure 5). None of the SL cDNAs detected sequences that comigrate with any genomic 1.4 kb SL repeat units, suggesting they are not closely linked to the tandem array or the major SL orphons (not shown).

Cloning of Genes Encoding SL cDNAs

Although no evidence of close linkage between genomic SL sequences and the genes detected by SL cDNAs was observed in Southern analyses, it might be difficult to detect a single SL dispersed from the tandem array under the conditions employed. We therefore cloned genomic sequences corresponding to four of the SL-containing cDNAs, previously determined to detect single-copy genes. VAT 5 DNA was partially digested with Bam HI or Mbo I, fragments ranging from 12 to 20 kb were cloned into bacteriophage λ 1059, and recombinants that hybridized with the SL cDNA probes were isolated. None of these clones hybridized with more than one cDNA, indicating that the genes examined are not closely linked. None of the genomic clones hybridized with the SL, or with the genomic 1.4 kb SL repeat unit even under relaxed conditions (3× SSC, 50°C). For example, in Figure 4, cDNA clone pSLc1 (lane C), and genomic DNA clone λ(SL)g1 (lane G) were digested with Cla I and hybridized either to a 5′ 250 bp subclone of pSLc1 (Figure 4A) or to the SL probe (Figure 4B). Both the cDNA and the genomic clones hybridize with the 5′ 250 bp fragment, but only the cDNA hybridizes with the SL.

The absence of the SL in these recombinants could be the result of cloning only the 3′ portion of the gene. To rule out this possibility, representative genomic clones of each of the four genes were analyzed by restriction enzyme mapping. In each case, clones were found that contained several (3–12) kilobases of DNA upstream from the structural gene (Figure 5). The genomic clone discussed above and shown in Figure 4 contained 8 kb 5′ to the structural gene. The restriction sites present in each cDNA, up to but not including the Xmn I site contained in

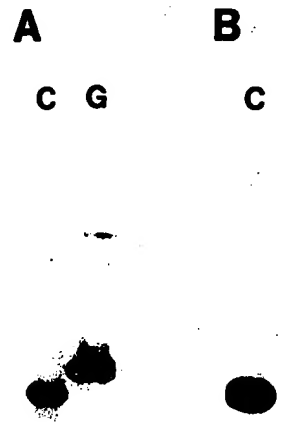

Figure 4. Hybridization of the SL Probe to Genomic and cDNA Clones

Genomic clones detected by pSLc1 were isolated. Genomic clone λ(SL)g1-3 (lane G) and cDNA clone pSLc1 (lane C) were cleaved with Cla I. Cla I has recognition sites within both of the vectors, as well as within the cloned structural gene (see Figure 5A). (A) Hybridization with a 250 bp 5′ fragment of pSLc1. (B) Hybridization with the SL probe.

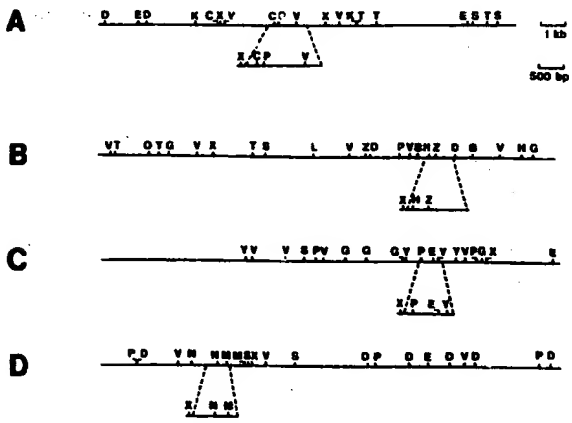

Figure 5. The Genomic Environment of Selected Genes That Use the SL

In each case, the top line shows the restriction map derived from overlapping genomic clones, while the lower line shows the map of the corresponding cDNAs. In some cases only the restriction sites closest to the gene are shown. Note that the Xmn I site (X) present in the SL was not observed in the corresponding position in the genomic clones. Restriction enzyme sites are marked as follows: Bgl I, Z; Bgl II, G; Bam HI, B; Cla I, C; Eco RI, E; Hinc II, H; Hind III, D; Kpn I, K; Mlu I, M; Nco I, N; Pst I, P; Pvu II, V; Sal I, L; Sma I, S; Sph I, Y; Sst I, T; Xho I, O; Xmn I, X.

(A) pSLc1 and its genomic environment; (B) pSLc2 and its genomic environment; (C) pSLc3 and its genomic environment; (D) pSLc4 and its genomic environment.

the SL (see Figures 2 and 5), were also present in the homologous genomic clones. For example, the Hinc II site, located 37 bp from the 3′ end of th SL in pSLc2, and the Pst I site, located 28 bp from the SL in pSLc3, are each found in the corresponding λ(SL)g2 and λ(SL)g3 genomic clones. In both cases, these sites are in the 5′ untranslated regions of the molecule. Thus the 3′ portion of the leader sequence is derived from sequences abutting the structural gene, while the 5′ portion containing the SL is transcribed from elsewhere in the genome.

Discussion

The experiments described here demonstrate that the 35 nucleotide SL is not unique to VSG mRNAs but is also present on many other trypanosome RNAs. Estimates based on the frequency of SL-containing clones in cDNA libraries suggest that a substantial fraction of cytoplasmic mRNAs may carry this 5′ sequence. Sequences common to many different cellular mRNAs have been described in other eucaryotic systems, but they are not 5′-terminal. For example, a 3′-terminal "suffix" sequence is found on approximately 2% of all Drosophila mRNAs (Tchurikov et al., 1982). Similarly the 3′ "Set 1" sequence is highly represented on mouse mRNAs transcribed during the first half of embryogenesis (Murphy et al., 1983). Another sequence is found on the intron regions of many RNA species in the rat—the "ID" sequences (Milner et al., 1984). This sequence appears to be a marker for brain-specific transcripts (Sutcliffe et al., 1984).

Each of these elements are repeated in their respective genomes, but unlike the tandemly repeated SL sequences, they are dispersed. They are encoded adjacent to or within the structural genes on whose transcripts they are found, again contrasting with the trypanosome SL, which is not detectably linked to the structural genes. The differences between the SL and other repetitive elements found on multiple transcripts, in particular regarding their locations in the genome and on the transcripts, indicate that different molecular mechanisms are responsible for their ubiquity.

Perhaps more similar to the trypanosome SL is the case of the coronavirus leader. Coronavirus RNAs are actually a series of nested transcripts that are 3′ coterminal and extend various distances 5′. Each RNA, however, shares the same 5′-terminal leader sequence (Spaan et al., 1983).

What mechanism accounts for the presence of the SL on trypanosome RNAs? Figure 6 depicts several alternative arrangements of SL and structural gene exons, and their resulting primary transcripts. Sequences just upstream from the SL resemble concensus sequences of eucaryotic RNA polymerase II promoters. The large tandem array of SL repeat units has therefore been hypothesized to provide multiple promoters for frequent initiation of VSG gene transcription (DeLange et al., 1983). However, the regulation of the transcription of the other genes whose transcripts contain the SL differs from that of VSG genes; those studied here are transcribed during both bloodstream and procyclic stages, and appear to be much less abundant than VSG mRNAs. Could all of these genes, with such different characteristics, be using the same promoter (see Figure 6A)? Recall that as with VSG genes, genomic sequences encoding the SL were not found within 3–12 kb of the structural genes encoding the SL-containing RNAs. Nor do any of these genes appear to be linked to one another. Thus if there is a single promoter used by all

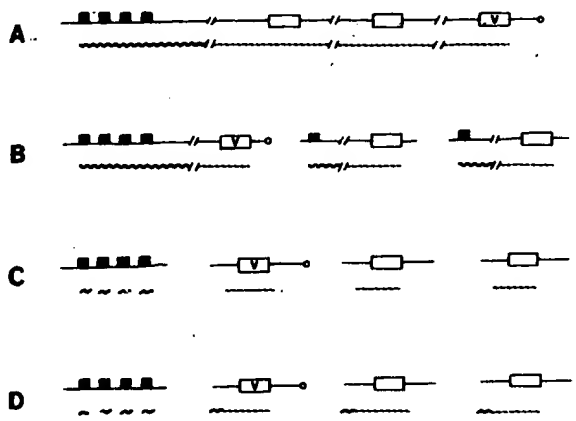

Figure 6. Models of SL and Structural Gene Transcription

In each case, the upper line depicts the genomic configuration and the lower line shows the primary transcription products. The large boxes indicate structural genes, with that labeled "V" being the active VSG gene. As indicated by the line terminating in a circle, the active VSG gene lies adjacent to a telomere. The small shaded boxes are the genomic SL sequences, either in the tandem array or in the orphons.

(A) One continuous transcription unit for all genes that use the SL. (B) Separate transcription units for each structural gene that uses the SL. (C) Discontinuous transcription, subsequent RNA joining. (D) Discontinuous transcription, SL RNA used as primer.

these genes, the primary transcription unit must be hundreds of kilobases long, provide for stage-specific and abundant transcription of the distal VSG gene (transcribed VSG genes are adjacent to a telomere), and moderate transcription of many other genes throughout the trypanosome life cycle.

Another possibility is that each of these genes has its own SL promoter, located far upstream (Figure 6B). Although each of the SL repeat units appear very similar, we have detected a low level of sequence heterogeneity between repeat units (unpublished results), which might be hypothesized to allow for differences in the regulation of transcription. For example, the developmentally regulated VSG gene could be colinear with (albeit distant from) the major array of repeat units and the three to four SL orphons of our T. brucei stock could serve as promoters for other SL genes. However, there are many more than four non-VSG genes that employ the SL sequence. Even if there were several unidentified orphons or several tandem arrays, there do not appear to be sufficient unlinked SL transcription units for each gene to have its own. A combination of the two hypotheses, with several large transcription units, one colinear with developmentally regulated genes, and one colinear with constitutively expressed genes, could also be proposed.

We believe it more likely that the SL repeat units are not linked to the structural genes, and that transcription is discontinuous. Two different hypotheses could be proposed. First, that the SL and structural gene sequences could be transcribed from separate DNA molecules, with the two RNA molecules then joined by intermolecular splicing or ligation (Figure 6C). Alternatively, the SL transcript could aid in the initiation of structural gene transcription (Figure 6D) as host RNA fragments serve as primers for transcription of influenza virus genes (Plotch et al., 1981). Interestingly, this reaction does not require homology between primer and virus sequences. The coronavirus leader sequence has also been postulated to function as a primer in discontinuous transcription (Spaan et al., 1983). With regard to these latter hypotheses, one would predict the presence of small SL-containing RNAs. We have recently characterized an RNA species of approximately 135 bp that hybridizes to the SL probe (Milhausen et al., submitted).

It is clear that the role of the SL is not limited to VSG gene expression in African trypanosomes—it is found in other trypanosomatids that do not undergo antigenic variation (Nelson et al., 1984), and on RNAs derived from genes whose pattern of expression differs dramatically from that of VSG genes. Although as yet no structural genes, aside from those encoding VSGs or tubulins, have been studied in trypanosomes, identification of the function of the genes employing the SL may serve to clarify its role in trypanosomatid gene expression. Alternatively, the SL may be present on virtually all trypanosome mRNAs, resulting from a requisite role in the transcription or RNA-processing machinery of the cell. The possibility of a unique process critical to gene expression in these pathogenic organisms may provide a target for rational chemotherapy of trypanosomiasis.

Experimental Procedures

Trypanosomes

The IsTaR serodeme of Trypanosoma brucei brucei was employed for all studies (Stuart et al., 1984). Cells of VAT 5 were converted to procyclic forms by in vitro culture (Hanas et al., 1975). DNA for Southern analyses and RNA for construction of cDNA libraries were isolated from VAT 5 cells and from procyclic cells cultivated for at least 2 months (Parsons et al., 1983; Milhausen et al., 1983). Total RNA isolated from VAT 3 cells or procyclic cells derived from VAT 5 was a gift of Dr. Jean Feagin.

cDNA and Genomic Clones

The construction of cDNA libraries in pBr322 has been described (Parsons et al., 1983). The generation of a library of genomic DNA (from VAT 1.5 cells) in bacteriophage λ 1059 is described by Aline et al. (submitted). SL-containing cDNAs were detected by hybridization with ^{32}P-labeled synthetic probe complementary to 22 nucleotides of the 35 nucleotide SL (see Figure 2) using hybridization conditions previously determined to detect specific SL sequences in the genome (see below).

Those cDNAs used for studies reported here were pTbSLc1-1(B), pTbSLc2-1(B), pTbSLc3-1(P), pTbSLc4-1(P), and pTbSLc5-1(P) and for convenience are designated in the text as pSLc1 through pSLc5, respectively. pSLc1 and pSLc2 were isolated from the bloodstream-stage cDNA library while pSLc3, pSLc4, and pSLc5 were isolated from the procyclic-stage library. Genomic clones corresponding to pSLc1 are designated λ(SL)g1-n (where n is the clone number); those corresponding to pSLc2 are designated λ(SL)g2-n, etc. Thirty-four λ(SL)g1 clones, four λ(SL)g2 clones, ten λ(SL)g3 clones, and ten λ(SL)g4 clones were isolated.

The portions of selected cDNAs that hybridized to the synthetic probe were subcloned into M13 vectors mp 8, mp 9, mp 18, or mp 19. The sequence was determined by the dideoxy chain termination method (Sanger et al., 1977).

Hybridization Analyses

Total RNA was fractionated by electrophoresis in 1.4% agarose-formaldehyde gels and transferred to nitrocellulose membranes (Milhausen et al., 1983). Hybridizations were in 5× SSPE (1× SSPE is 180 mM NaCl, 10 mM

NaH_2PO_4, 1 mM EDTA, pH 7.4), 0.2% sarkosyl, and 200 μg/ml denatured salmon testis DNA for 16–30 hr at 65°C (for cDNA probes) or 37°C (for the 22-mer probe). The final stringency of post-hybridization washes was 0.3× SSC at 65°C for cDNA probes and 5× SSC at 37°C for the SL probe (1× SSC is 150 mM NaCl, 15 mM NaCitrate, pH 7.0). Southern analyses were performed as previously described (Parsons et al., 1983). The final stringencies for post-hybridization washes were the same as those described above, except for VSG cDNA hybridization, which employed a 0.1× SSC, 65°C final wash. cDNAs were labeled with ^{32}P by nick translation to a specific activity of approximately 10^8 cpm/μg. The 22-mer, a gift of Dr. Phillip Barr (Chiron Corporation), was labeled at its 5′ end to a specific activity of approximately 5×10^6 cpm per pmole as previously described (Nelson et al., 1983).

Acknowledgments

The authors would like to thank Dr. Kenneth Stuart for providing the VAT 5 genomic library, Susan Sather for helpful discussions, and Cyndy Gartside for secretarial assistance. This work was supported in part by grants from the Rockefeller Great Neglected Disease Program and from the National Institutes of Health (AI 17309). N. A. is a Burroughs-Wellcome Scholar in Molecular Parasitology.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received May 9, 1984

References

Agabian, N., Thomashow, L., Milhausen, M., and Stuart, K. (1980). Structural analysis of variant and invariant genes in trypanosomes. Am. J. Trop. Med. Hyg. *29* (5) Suppl. 1043–1049.

Bernards, A., Van der Ploeg, L. H. T., Frasch, A. C. C., Borst, P., Boothroyd, J. C., Coleman, S., and Cross, G. A. M. (1981). Activation of trypanosome surface glycoprotein genes involves a duplication-transposition leading to an altered 3′ end. Cell *27*, 497–505.

Bernards, A., Michels, P., Lincke, C., and Borst, P. (1983). Growth of chromosome ends in multiplying trypanosomes. Nature *303*, 592–597.

Boothroyd, J., and Cross, G. (1982). Transcripts coding for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene *20*, 281–289.

Buck, G., Longacre, S., Raibaud, A., Hibner, U., Giroud, C., Baltz, T., Baltz, D., and Eisen, H. (1984). Stability of expression-linked surface antigen gene in *Trypanosoma equiperdum*. Nature *307*, 563–566.

Childs, G., Maxson, R., Cohn, R. H., and Kedes, L. (1981). Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. Cell *23*, 651–663.

DeLange, T., and Borst, P. (1982). Genomic environment of the expression linked extra copies for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. Nature *299*, 451–453.

DeLange, T. Liu, A. Y. C., Van der Ploeg, L. H. T., Borst, P., Trompe, M. C., and van Boom, J. H. (1983). Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes: a multiple promoter for VSG gene transcription? Cell *34*, 891–900.

Hanas, J., Linder, G., and Stuart, K. (1975). Mitochondrial and cytoplasmic ribosomes and their activity in blood and culture forms of *Trypanosoma brucei*. J. Cell Biol. *65*, 103–111.

Hoeijmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P., and Cross, G. A. M. (1980a). Novel expression-linked copies of the genes for variant surface antigens in trypanosomes. Nature *284*, 78–80.

Hoeijmakers, J. H. J., Borst, P., Van den Burg, J., Weissmann, C., and Cross, G. A. M. (1980b). The isolation of plasmids containing DNA complementary to messenger RNA for variant surface glycoproteins of *Trypanosoma brucei*. Gene *8*, 391–417.

Longacre, S., Hibner, U., Raibaud, A., Eisen, H., Baltz, T., Giroud, C., and Baltz, T. (1983). DNA rearrangements and antigenic variation in *Trypanosoma equiperdum*: multiple expression-linked sites in independent isolates of trypanosomes expressing the same antigen. Mol. Cell. Biol. *3*, 399–409.

Michiels, F., Matthyssens, G., Kronenburger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Cravador, A., Steinert, M., and Hamers, R. (1983). Gene activation and re-expression of a *Trypanosoma brucei* variant surface glycoprotein. EMBO J. *2*, 1185–1192.

Milhausen, M., Nelson, R., Parsons, M., Newport, G., Stuart, K., and Agabian, N. (1983). Molecular characterization of initial variants from the IsTat 1 serodeme of *Trypanosoma brucei*. Mol. Biochem. Parasitol. *9*, 241–254.

Milner, R., Bloom, F., Lai, C., Lerner, R., and Sutcliffe, G. (1984). Brain-specific genes have identifier sequences in their introns. Proc. Nat. Acad. Sci. USA *81*, 713–717.

Murphy, D., Brickell, P. M., Latchman, D. S., Wilson, K., and Finby, P. W. J. (1983). Transcripts regulated during normal embryonic development and oncogenic transformation share a repetitive element. Cell *35*, 865–871.

Nelson, R. G., Parsons, M., Barr, P. J., Stuart, K., Selkirk, M., and Agabian, N. (1983). Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in Trypanosoma brucei. Cell *34*, 901–909.

Nelson, R., Parsons, M., Selkirk, M., Barr, P., and Agabian, N. (1984). Sequences homologous to variant antigen mRNA spliced leader in Trypanosomatidae which do not undergo antigenic variation. Nature *308*, 665–667.

Overath, P., Czichos, J., Stock, U., and Nonnengasser, C. (1983). Repression of glycoprotein synthesis and release of surface coat during transformation of *Trypanosoma brucei*. EMBO J. *2*, 1721–1729.

Parsons, M., Nelson, R. G., Newport, G., Milhausen, M., Stuart, K., and Agabian, N. (1983). Genomic organization of *Trypanosoma brucei* variant antigen gene families in sequential parasitemias. Mol. Biochem. Parasitol. *9*, 255–269.

Parsons, M., Nelson, R., and Agabian, N. (1984a). Antigenic variation in african trypanosomes: DNA rearrangements program immune evasion. Immunol. Today *5*, 43–50.

Parsons, M., Nelson, R., Stuart, K., and Agabian, N. (1984b). Variant antigen genes of *Trypanosoma brucei*: genomic alteration of a spliced leader orphon and retention of expression linked copies during differentiation. Proc. Nat. Acad. Sci. USA *81*, 684–688.

Pays, E., Delronche, M., Lheureux, M., Vervoort, T., Bloch, J., Gannon, F., and Steinert, M. (1980). Cloning and characterization of DNA sequences complementary to messenger ribonucleic acids coding for the synthesis of two surface antigens of *Trypanosoma brucei*. Nucl. Acids Res. *8*, 5965–5981.

Pays, E., Van Meirvenne, N., LeRay, D., and Steinert, M. (1981a). Gene duplication and transposition linked to antigenic variation in *Trypanosoma brucei*. Proc. Nat. Acad. Sci. USA *78*, 2673–2677.

Pays, E., Lheureux, M., and Steinert, M. (1981b). The expression-linked copy of surface antigen gene in *Trypanosoma* is probably the one transcribed. Nature *292*, 265–267.

Plotch, S. J., Bouloy, M., Ulmanen, I., and Krug, R. M. (1981). A unique cap ($m^7GpppXm$)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. Cell *23*, 847–858.

Sanger, F., Nicklen, S., and Coulsen, A. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA *74*, 5463–5467.

Spaan, W., Delius, H., Skinner, M., Armstrong, J., Rottier, P., Smeekens, S., van der Zeijst, B., and Siddell, S. (1983). Coronavirus mRNA synthesis involves fusion of non-contiguous sequences. EMBO J. *2*, 1839–1844.

Stuart, K., Gobright, E., Jenni, L., Milhausen, M., Thomashow, L., and Agabian, N. (1984). The IsTat serodeme of *T. brucei*: development of a new *T. b. brucei* serodeme and characterization of early variants. Parasitology, in press.

Sutcliffe, J., Milner, R., Gottesfeld, J., and Lerner, R. (1984). Identifier sequences are transcribed specifically in brain. Nature *308*, 237–241.

Tchurikov, N., Naumova, A., Zelentsova, E., and Georgiev, G. (1982). A cloned unique gene of *Drosophila melanogaster* containing a repetitive 3′

exon whose sequence is present at the 3′ ends of many different mRNAs. Cell *28*, 365–373.

Van der Ploeg, L. H. T., Valerio, D., DeLange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982a). An analysis of the cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. Nucl. Acids Res. *10*, 5903–5923.

Van der Ploeg, L., Bernards, A., Rijsewijk, F., and Borst, P. (1982b). Characterization of the DNA duplication-transposition that controls the expression of two genes for variant surface glycoproteins in *Trypanosoma brucei*. Nucl. Acids Res. *10*, 593–609.

Van der Ploeg, L., Liu, A., Michels, P., DeLange, T., Borst, P., Majumder, H., Weber, H., Veeneman, G., and Van Boom, J. (1982c). RNA splicing is required to make the messenger RNA for a variant surface antigen in trypanosomes. Nucl. Acids Res. *10*, 3591–3604.

Williams, R., Young, J., and Majiwa, P. (1982). Genomic environment of *T. brucei* VSG genes: presence of a minichromosome. Nature *299*, 417–421.

Young, J., Donelson, J., Majiwa, P., Shapiro, S., and Williams, R. (1982). Analysis of genomic rearrangements associated with two variable antigen genes of *Trypanosoma brucei*. Nucl. Acids Res. *10*, 803–819.

cite

QH301.J6771
to 6/30 mic

STIC-ILL

453419

From: Wessendorf,Teresa
Sent: Saturday, June 28, 2003 2:13 PM
To: STIC-ILL
Subject: 09/990091

Please forward:
1.Cell, 38, 309-316 (1984).
2. J. Cell.Evol. 108, 229-241 (1989).
Txs. Bill
T. Wessendorf
Art Unit 1639
Rm. 2B17
MailRm. 3B01
308-3967

The Scanning Model for Translation: An Update

Marilyn Kozak

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Abstract. The small (40S) subunit of eukaryotic ribosomes is believed to bind initially at the capped 5′-end of messenger RNA and then migrate, stopping at the first AUG codon in a favorable context for initiating translation. The first-AUG rule is not absolute, but there are rules for breaking the rule. Some anomalous observations that seemed to contradict the scanning mechanism now appear to be artifacts. A few genuine anomalies remain unexplained.

THE scanning mechanism for initiation of translation in eukaryotes was proposed 10 years ago (112). Supporting evidence has accumulated at a slower rate than one might have wished, but a trickle sustained over 10 yr forms a decent-sized pond. Some remarkable experiments now being carried out in yeasts are yielding important new insights about scanning and other aspects of initiation, and the power of the yeast system promises additional breakthroughs in the coming years. Thus, it seems a good time to review what we've learned so far about how eukaryotic ribosomes select a particular AUG codon as the start site for translation.

Evidence from Higher Eukaryotes

The scanning model states that the 40S ribosomal subunit (carrying Met-tRNA$_i^{met}$ and various initiation factors) binds initially at the 5′-end of mRNA and then migrates, stopping at the first AUG codon in a favorable context for initiating translation. The model thus posits that both position (proximity to the 5′-end) and context contribute to selection of the initiation site. The simplest evidence of the importance of position is that the "first-AUG rule" holds for some 90–95% of the hundreds of vertebrate mRNA sequences that have been analyzed (127). A detailed discussion of position effects, along with an explanation for the 5–10% deviation from the first-AUG rule, will follow after a brief introduction to the context requirements for initiation.

From a recent survey of 699 vertebrate mRNAs (127), GCCGCCA$\overset{G}{C}$CAUGG emerged as the consensus sequence for initiation in higher eukaryotes, confirming and extending previously noted trends (117, 120). Site-directed mutagenesis experiments in which the expression of a cloned preproinsulin gene was monitored in transfected COS cells have confirmed the importance of G^{+4} as well as each of the consensus nucleotides from position −1 through −6 (122, 126)[1]. The importance of context was demonstrated by targeting mutations to the vicinity of the AUG initiator codon for preproinsulin as well as by targeting an upstream, out-of-frame AUG codon; in the latter case the inhibitory effect of the upstream AUG codon (i.e., its ability to block the initiation of preproinsulin from a downstream site) increased as the surrounding nucleotides increasingly resembled the consensus sequence. The importance of a purine in position −3 was confirmed by the discovery of a type of thalassemia in which a change in sequence, from CACCAUG to CCCCAUG, drastically impairs initiation of translation of α-globin (167). Experiments confirming (66, 92, 192, 248, 255) and exploiting the effects of context on initiation (169, 181, 222, 263)[2] have recently been reported from other systems. A purine (usually A) in position −3 is the most highly conserved nucleotide in all eukaryotic mRNAs, including those of vertebrates, plants (81), and fungi (188); and a mutation in that position affects translation more profoundly than a point mutation anywhere else (122). Indeed, as long as there is a purine in position −3, deviations from the rest of the consensus sequence only marginally impair initiation. In the absence of a purine in position −3, however, G^{+4} is essential for efficient translation (122) and the contributions of other nearby nucleotides can be detected (126). For practical purposes, an initiator codon can usually be designated "strong" or "weak" by considering only positions −3 and +4; I shall follow that convention in the rest of this paper.

The scanning model predicts that proximity to the 5′-end determines which AUG codon (in a good context) actually functions as the initiator codon. The nearly simultaneous discoveries of the m7G cap (232) and of "silent" 3′-cistrons in many viral mRNAs (243) constituted the first strong evidence that eukaryotic ribosomes are somehow restricted to initiating near the 5′-end. Some of the early in vitro evidence for silent 3′-cistrons has recently been confirmed in vivo (70, 281). In lieu of a scanning mechanism to explain the 5′ restriction in viral mRNAs, one might have postulated that the primary and/or secondary structure around a particular

1. Numbering begins with the A of the AUG codon as position +1; nucleotides 5′ to that site are assigned negative numbers.

2. In reference 222, although improving the context around the AUG codon increased the number of transformants 10-fold, the yield of protein could not be elevated above a threshold that was set by inefficiency at some later step in expression.

© The Rockefeller University Press, 0021-9525/89/02/229/13 $2.00

AUG triplet (which just happened to be 5′ proximal) caused it to be the preferred initiation site. To rigorously test the importance of position, therefore, a plasmid with four identical copies of the preproinsulin "ribosome binding site" was constructed and tested (118); the outcome, exclusive use of the first site in the tandem array, strongly supports a scanning mechanism. The scanning model predicts that, upon introducing an adventitious upstream AUG codon (in a good context), initiation should shift to the upstream site. Thus, if an upstream AUG codon is in the wrong reading frame, it should, and does, depress the yield of protein from the normal site (10, 39, 70, 105, 144, 156, 198, 214, 282). If a strong AUG codon is introduced upstream from, and in the same reading frame as, the normal initiator codon, the result is a stretched polypeptide with an NH_2-terminal amino acid extension. This has been documented with laboratory constructs (87, 118, 132, 148, 228, 244) and the same principle operates with some cellular and viral genes that use two promoters for transcription. In the latter cases, the longer transcript has an extra AUG codon (in a good context) upstream from the start of the shorter transcript; ribosomes initiate exclusively at the first AUG codon in each form of mRNA, producing "long" and "short" versions of the encoded protein that often mediate different biological functions (130). (If the 5′-proximal AUG codon in the longer mRNA were in a less favorable context, both versions of the protein could be translated from a single mRNA, as described below.)

Besides the wealth of evidence relating to position effects, here is some of the other evidence that supports and defines the scanning mechanism: (*a*) That the 40S ribosomal subunit/factor complex can migrate was deduced from the properties of complexes formed between reovirus mRNAs and wheat germ ribosomes in the presence of edeine (131). Those observations were subsequently confirmed with extracts from mammalian cells (35). (*b*) Other in vitro experiments revealed that sliding requires ATP hydrolysis and that, in the absence of ATP, 40S subunits are trapped upstream from the AUG codon, as the scanning model predicts (115). (*c*) The inability of eukaryotic ribosomes to bind to circular mRNAs is strong evidence against direct binding at the AUG start site (109, 113). That site was identical in the linear and circular forms of each template, but only the linear form could bind to wheat germ or reticulocyte ribosomes. Control experiments showed that both linear and circular templates bound to bacterial ribosomes, emphasizing the fundamental difference between prokaryotes and eukaryotes in the mechanism of initiation (119). (*d*) Translation of various mRNAs is inhibited when a hairpin structure, or a DNA–RNA hybrid, or a hybrid with anti-sense RNA is introduced near the 5′-end of the mRNA in a way that does not occlude the AUG codon or the m7G cap (36, 75, 123). The simplest interpretation, albeit not yet verified by footprinting, is that 40S ribosomal subunits can still bind to such mRNAs but cannot migrate beyond the duplex barrier.

Evidence from Yeasts

Yeast mRNA sequences resemble those of beasts in that translation begins at the first AUG codon in 95% of the genes examined (29). The rare occurrence of upstream AUG codons is neither an accident, nor a problem, but a means to regulate the translation of some interesting genes (168, 250, 261, 277). Two yeast genes have been subjected to exhaustive genetic probing, the results of which strongly support a scanning mechanism. One set of experiments began with a CYC*1* allele in which the normal AUG initiator codon had been inactivated (235). From the structure of revertants that had regained cytochrome c function, it appeared as if an AUG codon introduced anywhere within a stretch of 37 nucleotides could function in initiation. Donahue and Cigan (44) used a complementary approach to reach the same basic conclusion. They began with a specially designed *HIS4* allele in which initiation at the 5′-proximal AUG codon gave a His^- phenotype, while initiation from the next AUG downstream restored gene expression. That unique genetic behavior enabled them to select for mutations that reduced or abolished ribosomal recognition of the first AUG codon. The only point mutations identified in the search were alterations in the AUG codon itself, elegantly confirming the importance of "position" and suggesting that, if flanking sequences affect recognition of the AUG codon in yeast, the effects must be subtle.

When the latter question was addressed directly, by mutating nucleotides in the vicinity of the AUG codon, only modest effects (twofold or less) were found (6, 31, 277, 287). Indeed, although yeast mRNA sequences almost always have an A in position −3, the rest of the vertebrate consensus sequence is not evident (29), suggesting a different role for (or different sensitivity to) context in the two systems. Even the conserved A^{-3} might simply reflect the overall A richness of 5′-noncoding sequences in lower eukaryotes. The significance of the A-rich, G-deficient leader sequences on mRNAs from yeasts and other lower eukaryotes (26, 48, 51, 104, 172, 208, 245) is unknown, but an inevitable consequence is that such leader sequences lack extensive secondary structure, which seems to be more of an impediment to translation in yeasts (6, 31) than in higher eukaryotes (123).

Other recent experiments from Donahue's laboratory have yielded one remarkable new insight and validated one critical old assumption about the mechanism of initiation of translation. This time Donahue's group began by inactivating the AUG initiator codon in the yeast *HIS4* gene. Their subsequent search for second-site suppressor mutations led them to eIF2, the protein factor that escorts Met-$tRNA_i^{met}$ onto the ribosome. They found that a mutation in the β subunit of eIF2 suppressed the His^- phenotype by allowing ribosomes to initiate at the first UUG codon in *HIS4* mRNA (45). Thus, eIF2 is an active participant (and the only protein factor so far implicated by genetic criteria!) in the mRNA binding-and-scanning step of initiation. The elegant follow-up experiment was to change the anticodon sequence from 3′-UAC-5′ to 3′-UCC-5′ in one of the $tRNA_i^{met}$ genes of *S. cerevisiae*; the mutant form of $tRNA_i^{met}$ directed ribosomes to initiate at (the first) AGG instead of the usual AUG codon (30). This is direct proof that the initiator codon is recognized primarily by base pairing with the anticodon in Met-$tRNA_i^{met}$, as had always been (only) supposed. And the experiment is compelling proof of scanning.

Explicable Exceptions to the First-AUG Rule

There are a number of well-characterized viral and cellular mRNAs in which translation is not limited to the AUG codon nearest the 5′-end, but even these "exceptional" mRNAs adhere to rules that are consistent with a scanning mechanism. For example, although initiation is not restricted to the first

AUG codon in the examples discussed below, initiation is at least limited to AUG codons in the vicinity of the 5′-end. In no case do eukaryotic ribosomes initiate de novo in the middle of an mRNA. In several cases in which one or more small open reading frames (ORFs)[3] precede the major ORF, the small upstream ORFs are translated (65, 76, 91, 101, 277). The "problem," therefore, is not that the 5′-proximal AUG codon is missed but that it is not used exclusively.

Initiation at downstream AUG codons occurs, not haphazardly, but under three specific conditions: (*a*) When there are fewer than 10 nucleotides between the cap and the first AUG codon, ribosomes may initiate at the first and second AUG codons (99, 199, 251). This rule, deduced from the behavior of natural mRNAs, has not yet been verified by systematic experiments; it is supported, however, by some results of manipulating late leader sequences on SV-40 mRNAs (39, 72). (*b*) When the first AUG codon lies in an unfavorable context for initiation (i.e., when position −3 is C or U; or when position −3 is G and position +4 is not G), "leaky scanning" enables some ribosomes to reach and initiate at the second AUG codon. The hypothesis (125) that leaky scanning underlies the ability of bifunctional viral mRNAs to direct the synthesis of two separately initiated proteins is supported by the effects of mutations on the translation of two such mRNAs. Thus, Sedman and Mertz (228) probed the translation of SV-40 late 19S mRNA by introducing mutations near the 5′-end, and found that the relative yields of VP2 and VP3 (the second initiated downstream from the first) varied in accordance with the scanning rules. Mutagenesis of the Sendai virus P/C gene (where proteins P and C are translated from one mRNA in different, overlapping reading frames) also gave results consistent with leaky scanning, with the interesting twist that initiation occurs from three sites in that message: the first ACG codon functions because it lies in an excellent context (GCCACGG), but functions poorly because it is not AUG; the next AUG codon in a rather poor context (CGCAUGG) functions inefficiently; thus most 40S subunits advance to and initiate at the third start site, AAGAUGC (37, 74). Thus far, 17 viral mRNAs (reviewed in reference 124) have been shown to produce two, or rarely three, overlapping proteins by initiating at a weak upstream, as well as the next downstream, AUG codon. An important addition to the list of bifunctional mRNAs is the pX transcript of human T-cell leukemia virus type I, which directs synthesis of both p27 (from the first, weak AUG codon) and p40 (173, 241). The recently determined sequences of turnip yellow mosaic virus genomic RNA (165) and one of the simian rotavirus SA11 gene segments (164) suggest that they, too, should produce two proteins by leaky scanning. The only viral mRNA known to produce two proteins from overlapping reading frames, the first of which initiates with an anomalously strong AUG codon, is the NA/NB mRNA of influenza B virus (234). The explanation in that case might involve slippage on the run of A residues flanking the first AUG codon. (*c*) The third condition that allows access to internal AUG codons is reinitiation. When an AUG codon upstream from the start of the major protein coding sequence lies in a favorable context, thereby precluding leaky scanning, ribosomes can still reach the downstream initiation site provided that a terminator codon occurs in-frame with the first AUG codon and upstream from the second (121, 128, 148). In such cases the first (invariably small) ORF is translated (65, 76, 91, 101, 277), after which 40S ribosomal subunits apparently resume scanning and reinitiate farther downstream. (If initiation factors dissociate from the 40S subunit slowly rather than instantaneously, upon the addition of a 60S subunit and commencement of peptide bond formation, reinitiation might be possible after the translation of an upstream "minicistron," as has been observed, but not after the translation of a full-sized cistron. That as-yet-untested hypothesis would explain why 3′-cistrons in viral mRNAs are usually silent [243].) The influence of flanking sequences on AUG codon recognition follows the same hierarchy in the reinitiation mode as in the primary scanning mode (122), but not every downstream AUG codon in a good context is an efficient site for reinitiation. The efficiency of reinitiation at a downstream AUG codon steadily improves, for example, as its distance from the upstream ORF increases. Consequently, reinitiation is not limited to the first strong AUG codon after the 5′-proximal ORF. (See reference 128 for evidence and further discussion of this.) Because reinitiation is usually inefficient with natural mRNAs, the presence of short upstream ORFs usually reduces translation of the downstream ORF (69, 72, 102, 185), albeit not as severely as were there no terminator codon between the upstream AUG codon and the downstream ORF. Failure to consider the contribution of leaky scanning probably explains some reports (193, 262) in which the apparent requirements for reinitiation differed from what I have described here.

Some Harder Cases

The popular press (88) has recently announced that picornaviruses "break the rules" by allowing a 40S ribosomal subunit to bind directly to an internal site (somewhere upstream from the start of the major ORF) in lieu of the usual end-dependent mode of entry. The key experiments (196) involve the translation of a dicistronic transcript of the form TK-PV-CAT, where the 736-nucleotide poliovirus leader sequence (PV) separates the 5′-proximal thymidine kinase gene (TK) from the 3′-proximal chloramphenicol acetyltransferase gene (CAT). COS cells transfected with the dicistronic vector clearly produced some CAT protein, but that result hardly warrants the conclusion that the presence of the poliovirus leader sequence allows the efficient translation of CAT by direct internal initiation. One problem is that efficiency was claimed without having been demonstrated; i.e., it was not shown how much CAT protein was produced in vivo from the dicistronic vector relative to a monocistronic CAT transcript that bears a normal leader sequence.[4] The most serious deficiency, however, is that the Northern blot which was offered as proof that the dicistronic transcript is the only form of CAT mRNA in transfected cells was much too faint to prove the point. When translation of dicistronic mRNAs was studied in vitro, on the other hand, the unbound mRNA pool was found to be completely degraded after only 10 min of incubation, which makes it hard to believe that many, if any, transcripts were intact after 60 min, when the CAT yield was

3. *Abbreviation used in this paper*: ORF, open reading frame.

4. Indeed, the poliovirus leader sequence does not seem to support efficient translation even when it is at the 5′-end of a transcript. See the change in scale in Fig. 2 of reference 266.

measured. In the absence of rigorous proof that dicistronic mRNAs are the only transcripts available for translation, the recent picornavirus experiments (90; 196) are no different from older in vitro experiments (see below) in which the appearance of internal initiation has turned out to be an artifact. That some dicistronic mRNA was associated with polysomes (196) is not surprising, since the 5′-proximal TK cistron would be functional. To prove that the 3′-proximal CAT cistron could also function, the authors should have immunoprecipitated polysomes engaged in CAT synthesis and asked whether all (or any!) of the mRNA thus selected was dicistronic. In capable hands such experiments can yield answers (8). In a different approach, Pelletier and Sonenberg (196) showed that their dicistronic transcript formed "disomes" in the presence of sparsomycin, from which they assumed, without direct evidence, that one of the two ribosomes was bound to the PV sequence in the middle of the transcript. Years ago similar ideas were entertained with other viral mRNAs that formed disomes under conditions of initiation (201); but, in every instance where the ribosome-protected sites were sequenced, both sites mapped to the 5′-noncoding sequence (2, 202, 267). Thus the mere formation of disomes in sparsomycin-inhibited extracts is not evidence of a functional 3′-cistron. In view of all these shortcomings, all we yet know about the translation of poliovirus mRNA is that it occurs in the absence of the usual m7G cap and despite the burden of eight (mostly weak) upstream AUG codons (207).

Turning from viral to cellular mRNAs, the 5–10% of vertebrate mRNAs[5] that have upstream AUG codons are an interesting, nonrandom set. Many protooncogenes and growth-control genes produce mRNAs with upstream AUG codons (41, 68, 189, 215, 273, 283, 285)[6] which may be lost during the rearrangements that accompany activation (156 and references therein). The cDNAs that have been characterized from homeobox genes are also likely to have upstream AUG codons (24, 239), although it is not clear that those cDNAs correspond to functional mRNAs. The transcript that corresponds to one AUG-burdened homeobox cDNA, for example, is restricted to the nucleus (170). Cellular genes that produce mRNAs with upstream AUG codons often use alternative promoters and/or splice sites to generate supplementary transcripts in which the leader sequences are less problematical (68, 155, 179, 197, 210, 212, 223, 259, 269). Indeed, the 5′ variability is sometimes so extensive that no two cDNAs from a given gene have the same 5′ noncoding sequence (145, 155)! This underscores the need to distinguish between functional and nonfunctional (or minimally functional) cDNAs and mRNAs, a difficult problem that only a few investigators have tackled (3, 204). It might be mentioned in passing that, among growth control as well as housekeeping genes, G–C-rich leader sequences are much more pervasive, and thus more of a potential problem, than upstream AUG codons (23, 53, 77, 89, 134, 136, 138, 163, 166, 178, 191, 229, 231, 252, 270, 280). In the case of the c-*sis*/PDGF-2 mRNA, which has both upstream AUG codons and a G–C-rich leader sequence, it is primarily the formation of secondary structure (implied by the G–C richness) that restricts translation in vivo (209).[7] This is noteworthy because the inhibitory effects of secondary structure may be susceptible to environmental modulation (129) while there is, as yet, no evidence that the inhibitory effects of upstream AUG codons are regulatable in vertebrates, as they are in the GCN4 gene of yeast (86, 220, 267a). In vertebrates, the solution to upstream AUG codons is to get rid of them (68, 155, 179, 197, 210, 212, 223, 259, 269).

Lessons from cDNA Irregularities

Atypical or erroneous cDNA sequences have sometimes been mistaken as evidence against the generality of the scanning mechanism. Among cellular mRNAs that were initially reported to contain a slew of upstream AUG codons, many of the worst offenders have been exonerated as more data has emerged. For example, some long, AUG-burdened leader sequences have been recognized belatedly as errors in cloning or sequencing (58, 97, 183, 268; 52 corrected in 286; 59 corrected in 194; 110 corected in 153; 5 corrected in 56; 34 corrected in 103; 78 corrected in 139).[6] In some cases the error was simply that the cloned cDNA did not include the entire coding sequence: an internal AUG codon was mistakenly assumed to be the site of initiation, and therefore several upstream AUG codons, actually part of the coding sequence, were thought to burden the 5′-noncoding sequence (9 corrected in 79; 21 corrected in 22; 28 corrected in 94; 141 corrected in 142; 157 corrected in 254; 176 corrected in 175). The spurious upstream AUG codons in some cDNA sequences reflect derivation of the cDNA from a minor mRNA species that has an atypically long leader; the bulk of the transcripts from the same gene were shown to have a much shorter leader sequence, and no upstream AUG codons (95, 111, 151 corrected in 71).[6] As cloning efforts have gradually turned toward complicated regulatory genes, it has become fairly common to find cDNAs that correspond to incompletely processed transcripts (42, 63, 96, 146, 216, 225, 257). Thus, there are cases in which a bevy of out-of-frame AUG codons near the 5′-end of a cDNA sequence actually reside in an intron, which is not present, of course, in the functional mRNA (230, 236, 247, 272, 147 corrected in 64). For some *Drosophila* transcripts the excision of a 5′-intron, and consequent activation of translation, are developmentally regulated (17). If regulated (or merely inefficient) splicing is more widespread than we realize at present, most of the still-problematical cDNA sequences with multiple upstream AUG codons (e.g., 57, 83, 108, 177, 186, 206, 256) might eventually be traced to intron-containing pre-mRNAs. There are some mammalian cDNA sequences with AUG-burdened leader sequences that have not yet been formally recognized as introns, but that possibility is consistent with their localization in the nucleus (20, 63, 170); or their inability to be translated unless the upstream AUG codons are removed (171, 174, 187); or the presence of a typical 3′ splice junction motif at the point of divergence between two cDNA sequences (compare 154 with 260, 224 with 159, 135 with 152, and cDNAs *a* and *b* in reference 149); or the fact that an intron

5. This number, which comes from reference 127, is probably inflated for reasons described in the next section of the text.

6. See reference 127 for additional documentation.

7. In other cases (47, 107), the regulatory effect of leader sequences was demonstrated by deletion mutagenesis rather than site-directed mutagenesis; thus it is not clear whether an upstream AUG codon or the potential for secondary structure accounts for the observed inefficient translation of the wild-type sequence.

interrupts the 5′-noncoding sequence in other members of the gene family (compare 133 with 18, and 205 with 19). Many of the anomalous deductions about upstream AUG codons in mRNA sequences are favored, unfortunately, by the common practice of selecting and sequencing the longest cDNA clone! The main point here is not that cellular mRNA leader sequences never have upstream AUG codons but that such sequences are not nearly as common as superficial reading of the literature would suggest. When upstream AUG codons do occur they are a red flag, a clue to expect some sort of regulation (such as promoter switching or regulated splicing) or, at least, inefficient translation of the gene in question.

Artifacts of Cell-free Translation Systems

The results of some in vitro translation experiments have been taken as evidence against the generality of scanning. Five examples are often cited as evidence that eukaryotic ribosomes can bind directly to a site in the interior of a message without having traversed the upstream sequence. Rice et al. (213) have discussed the evidence for "internal initiation" in flaviviruses and have suggested credible alternative interpretations. As I cannot improve on their cogent discussion, I will simply recommend it to the reader and move on. For each of the other four examples, a synopsis of the arguments for and against internal initation follows.

(*a*) An SP6 vector-derived transcript corresponding to the adenovirus DNA polymerase gene directed the in vitro synthesis of the 120K (nearly full-sized) polymerase and a 62K polypeptide that mapped to the 3′-end of the pol coding sequence (80). Because cDNA corresponding to the 5′-portion of pol arrested the translation of the 120K but not the 62K product, the authors suggested that there is an independent ribosome-entry site at the midpoint of the pol ORF, and that the 62K protein was initiated internally. A less heretical explanation is that nuclease attack might have generated an mRNA fragment in which the initiator codon for 62K was near the 5′-end. The authors considered that explanation less likely because the translation results did not change when RNasin (an inhibitor of RNase) was included, and because ^{32}P-labeled mRNA was only slightly degraded during the first 15 min of incubation in the reticulocyte extract. In retrospect, however, either RNA cleavage or some other artifact must have occurred in the in vitro experiments, because the 62K protein was not synthesized in vivo from a plasmid that produced an abundance of functional 140K DNA polymerase (237).

(*b*) When vesicular stomatitis virus mRNA P was translated in a reticulocyte extract, the products were the full-sized P protein and a 7K COOH-terminal fragment thereof (84). The 7K polypeptide was attributed to internal initiation based on hybrid arrest experiments; i.e., a cDNA fragment complementary to the 5′-end of mRNA-P abolished translation of P but not the 7K protein. RNA cleavage was considered a less likely explanation because RNasin was present during the incubation and because the relative yield of 7K did not increase with time, as it might have were its template a degraded RNA. The second argument is weak, however, because one cleavage event might activate an mRNA fragment for 7K translation while the next clip, within the 7K coding sequence, might inactivate it. RNasin is not adequate insurance, as proven by the adenovirus story. The fact that cap analogues did not inhibit 7K translation (85) is equally consistent with internal initiation or initiation at the 5′-end of a broken template. A strong hint that the 7K protein is an artifact came from an experiment in which a transcript representing only the 3′-portion of the P gene (derived by subcloning that region into an SP6 vector) was translated in vitro: three small polypeptides were made, and the two "nonphysiological" products were more abundant than the "authentic" 7K band (84). Virus-infected cells contain barely detectable amounts of a protein that does not exactly comigrate with the 7K in vitro band (84) and might itself be a degradation product.

(*c*) In the case of infectious pancreatic necrosis virus, the claim (162) is that genomic RNA-A directs the independent synthesis of three major proteins: VP2, NS, and VP3. There is no absolute need to postulate three initiation events inasmuch as the sequence of genome segment A reveals a single ORF that encodes VP2-NS-VP3 as a fusion protein (49), and a protease activity that maps with NS is able to release the mature proteins from the common precursor (50). Thus the speculation is that, in addition to the polyprotein mode of translation, NS and VP3 can be translated by a second mechanism which involves internal initiation. The main supporting evidence seems to be that, during a short pulse with [^{35}S]methionine in vitro, VP2, NS, and VP3 acquired label simultaneously (162). But that observation is equally consistent with independent internal initiation of three proteins from one mRNA or translation of each protein from an independent template, generated by RNA degradation. In addition to VP2, NS, and VP3, in vitro translation of RNA-A produced countless other polypeptides that are never seen in vivo (50, 162), which is strong reason to suspect an in vitro artifact. The NH_2-terminal amino acid sequence of virion-derived VP3 has not yet been analyzed to ascertain the presence of either methionine, which would be consistent with internal initiation, or a protease-recognition sequence, which would implicate a polyprotein as the sole source of VP3 in vivo. Lacking such evidence, the argument for internal initiation is weak.

(*d*) The last example is poliovirus. When RNA from virions was translated in a reticulocyte lysate, the earliest detected products mapped to the 3′ portion of the genome (region P3), suggesting internal initiation site(s) (200). In support of that idea, *N*-formyl[^{35}S]methionine (a marker of initiation) was incorporated into polypeptides from region P3 (46). The question is whether internal start sites are accessible to ribosomes in intact poliovirus mRNA or whether RNA fragments are the functional templates. The anomalous internal sites detected in the reticulocyte system did not function in HeLa cell extracts, nor when an aliquot of HeLa cell extract was added to the reticulocyte lysate. The authors' interpretation was that HeLa cell extracts contain undefined factors that promote 5′ initiation, and "the deficiency (of those factors in the reticulocyte system) resulted in the ability of ribosomes to initiate translation on internal sequences" (46). One might think that deficiency of a required component would preclude translation, rather than endow ribosomes with a novel power; but that is debatable. The fact is that no virus-promoting initiation factor has yet been purified from HeLa cells, but HeLa extracts have been shown to contain an RNase inhibitor (249)! Finally, an experiment that might have detected internal initiation in vivo failed to do so. The experi-

ment involved poliovirus mutants in which small changes in the 5′ non-coding sequence depressed translation from the normal 5′ proximal site (265). Were there an independent initiation site in the interior of the poliovirus genome, it should have remained functional; but the residual level of translation in mutant-infected cells showed no enrichment for P3 products. Thus, although the mechanism of initiation at the 5′ proximal site in poliovirus mRNA remains unclear, the existence of an internal initiation site in the 3′ third of the genome can almost certainly be dismissed as an artifact.

Curran and Kolakofsky (38) have interpreted some recent results from the Sendai virus system as evidence that ribosomes bind to the 5′ end of the "P" transcript and jump some 1,500 nucleotides to the start of the "X" ORF, thus translating the X protein by a cap-dependent, scanning-independent process. Their hypothesis is based largely on the ability of cap analogues to inhibit X synthesis, but the hypothesis is contradicted by the virtual absence of X translation in vitro unless the P transcript is cleaved! In view of the many precedents for activation of internal initiation sites by mRNA cleavage (13, 116, 140, 195) and for inhibition of translation of uncapped transcripts by cap analogues (15, 61, 100, 227, 246), the interpretation offered by Curran and Kolakofsky seems unwarranted.

The main point here is not to debunk a few spurious claims against the generality of scanning but to illustrate the need for caution whenever cell-free systems are used to study translation. The tendency to see "internal initiation" more often in reticulocyte (12, 14, 40, 73, 180, 238, 264, 271, 279, 284) than in wheat germ translation systems might simply reflect the greater ease of translating broken (hence uncapped) mRNAs in the former system. A final caution is that initiation at non-AUG codons occurs far more efficiently in vitro than in vivo (4, 74, 192, 275) and therefore cell-free systems are not a reliable way to explore the rare, interesting situations in which eukaryotic ribosomes seem to initiate at codons other than AUG (10, 11, 37, 74). The usual failure of eukaryotic ribosomes to initiate at non-AUG codons in vivo is illustrated by the ability of point mutations in the AUG initiator codon to abolish gene expression (27, 32, 44, 72, 182, 203, 211, 217, 258). Conversely, bacterial genes that initiate with a GUG codon require the substitution of an AUG codon for successful expression in mammalian cells (137, 240).

Putting the Steps Together

Here is a brief statement of what we know and what we have yet to learn about the first three steps in initiation: binding of the 40S-ribosome/Met-tRNA/factor complex to mRNA; scanning; and recognition of the AUG initiator codon.

Step 1

The ubiquitous m7G cap and the associated cap-binding protein(s) (233) explain the predilection of eukaryotic ribosomes to engage mRNAs at the 5′-end. The initiation mechanism is end dependent even in those uncommon instances in which a cap is absent, however, prompting the idea that the 40S subunit might thread onto the 5′-end of mRNA (113). Microscopic data obtained with a new image processing technique indeed suggests the possibility of a channel running through the neck region of the 40S ribosomal subunit (60) which could be the needle's eye. (If the scanning mechanism were really a threading mechanism, the ability of 40S ribosomal subunits to hold on and reinitiate downstream might almost have been expected!) Apart from cap-recognition factor(s), we do not understand the precise function of any of the initiation factors that mediate early step(s) in the binding of mRNA to ribosomes. The remarkable finding (45) that a point mutation in eIF2 (the Met-tRNA$_i^{met}$-binding factor) affects where ribosomes initiate on mRNA illustrates the potential for surprises in this area.

Step 2

All we yet know about the mechanism of 40S subunit locomotion is that ATP is required (115), in common with other proteins that bind to nucleic acids and then (in an equally unknown manner) slide (25, 54). Scanning seems to be facilitated by something in the cytoplasm of mammalian cells, either a soluble protein or a 40S ribosome-associated activity, that has considerable ability to unwind duplex structures in the 5′-noncoding region of mRNA (123). The 40S subunit apparently migrates linearly, as evidenced by the ability of an upstream AUG codon to completely suppress initiation from a downstream site, even when the downstream AUG codon occurs in the same context as, and only five nucleotides beyond, the first AUG triplet (Fig. 7 in reference 121). The inability of ribosomes to "jump" a hairpin structure that is too stable to be melted is further evidence of the linearity of scanning (123). We do not know whether the 40S ribosome advances nucleotide-by-nucleotide or triplet-by-triplet; but, if it is the latter, each entering ribosome must pick at random any one of the three possible frames. If scanning were uniquely phased, a one- or two-nucleotide insertion between the cap and the AUG codon should drastically impair translation; in fact, such mutations are usually innocuous. An interesting possibility is that 40S ribosomal subunits might be nudged into the correct phase by the GCC or ACC motif that immediately precedes the AUG codon in vertebrate mRNAs. That would explain the ability of $^{A}_{G}$CC to enhance initiation when the purine occurs in position −3 or (less effectively) position −6, but not when the $^{A}_{G}$CC motif is shifted out-of-phase with respect to the AUG codon (122, 126). The observation that increasing the length of the leader sequence never impairs (and sometimes enhances) translation (129) suggests that, at least in the absence of secondary structure, scanning is not the rate limiting step in initiation.

Step 3

The migrating 40S ribosomal subunit stalls at the first AUG codon, which is recognized in large part by base pairing with the anticodon in Met-tRNA$_i^{met}$ (30). The stop-scanning step, and hence selection of the initiator codon, is susceptible to modulation, however: by context, at least in vertebrates (122, 126); by mutant forms of yeast eIF2 that compensate for (stabilize?) a weakened codon/anticodon interaction (45); by the antibiotic edeine (131); and by varying the concentration of Mg^{2+} and other ions in cell-free translation systems (114, 192, 221, 275).

Implications, Limitations, and Alternatives

In addition to correctly predicting the start sites for translation in the majority of viral and cellular mRNAs, the scanning model has informed the design of a variety of experi-

ments, including the construction of mammalian vectors that can initiate translation in all three reading frames (169) and the development of a "translational assay" for monitoring immunoglobulin gene rearrangements (55). The scanning mechanism explains the considerable effort expended by viruses (125) and some cells (93) to convert polycistronic transcripts into translatable (monocistronic) mRNAs; rationalizes an alternative strategy where several (up to seven!) cistrons are fused to generate some remarkable multifunctional proteins (62, 82, 98, 158); and justifies the observed inefficient translation of some critical genes, including some protooncogenes (156, 204, 209). The inhibitory effect of upstream AUG codons has been incorporated into an interesting model regarding the mechanism of allelic exclusion (143).

The scanning model is compatible with various modes of negative regulation, such as inhibition by upstream AUG codons (86, 267a), or by a repressor protein (219), or by secondary structure within the 5′-noncoding region; but the scanning mechanism seems incompatible with intricate schemes for enhancing the translation of specific mRNAs. Indeed, apart from the ubiquitous m7G cap and the conserved sequence around the initiator codon, the only other defined feature that has been shown to increase translational efficiency in eukaryotes is a long, unstructured 5′-sequence (129). Although leader-shuffling experiments have documented the ability of certain cellular and viral 5′-sequences to support efficient translation (16, 67, 106, 150, 253, 278), the application of mutagenesis techniques has consistently failed to pinpoint an effector motif within such leader sequences (160, 242). Thus, the data so far support the prediction (125) that, beyond the m7G cap and a favorable context near the AUG codon, what makes a "good" 5′-noncoding sequence is merely the absence of any unfavorable features.

An alternative to the three-step initiation mechanism described above is that 40S ribosomal subunits might bind directly to an internal site in some rare mRNAs. Although no credible evidence yet supports that possibility, it can never be ruled out. Neither will it ever become a fact unless the deficiencies in current experiments are recognized. One cannot ignore mRNA degradation in vitro as if it were a minor irritation. In vivo experiments have other limitations. The most carefully executed Northern assays sometimes fail to detect transcripts which, genetic evidence tells us, must be present and functional in cells (10). When it comes to reporter genes, the more sensitive the assay for protein expression, the harder it is to pinpoint the functional mRNA. Thus, a recent report of supposedly efficient expression of adenosine deaminase from the 3′-end of a dicistronic vector (161) was interpreted more cautiously when the high specific activity of adenosine deaminase was taken into account (184).

The existence of credible rules for initiation calls attention to occasional mRNAs that break the rules. In fact, very few viral mRNAs remain incompletely understood. It seemed for a while as if poxviruses might follow different rules for translation inasmuch as, in nearly all of the viral late genes, the ATG initiator codon is preceded at the DNA level by a T in position −3 (218, 276); we now know, however, that a complicated transcriptional mechanism replaces the undesirable T with the preferred A in position −3 (190, 226). I still cannot explain the translation of picornaviruses, cauliflower mosaic virus (43), or the putative bifunctional mRNAs from Epstein Barr virus (274) and cottontail rabbit papillomavirus (7). In the last two cases, however, the transcription patterns are complex and it seems possible that the downstream protein is translated, not from the 3′-end of the recognized bicistronic mRNA, but from a scarce monocistronic mRNA that has not yet been detected. One way to rationalize the peculiar leader sequences on retrovirus and poliovirus mRNAs is to suppose that features that compromise translation are tolerated because those very features are required to replicate (33) and package the viral genome (1). Moreover, viruses can and do compensate for inefficient translation by using their efficient transcription signals to flood the cell with mRNA (125). As for the structure of cellular genes, I have argued herein that the presence of numerous upstream AUG codons in a cDNA sequence constitutes a strong hint that the cDNA might represent an intron-containing pre-mRNA, rather than the funtional mRNA from the gene in question. The growing number of such sequences in the vertebrate cDNA catalogue raises the interesting possibility that the final, regulated step in the expression of many critical genes is the conversion of a stable, untranslatable precursor to a functional mRNA.

Research in the author's laboratory is supported by the National Institutes of Health (grant GM-33915).

Received for publication 15 November 1988.

References

1. Adam, M. A., and A. D. Miller. 1988. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J. Virol.* 62:3802–3806.
2. Ahlquist, P., R. Dasgupta, D. S. Shih, D. Zimmern, and P. Kaesberg. 1979. Two-step binding of eukaryotic ribosomes to brome mosaic virus RNA3. *Nature (Lond.).* 281:277–282.
3. Amrein, K., K. Dorsch-Hasler, B. Lutz, and M. Lezzi. 1988. Two transcripts of the same ecdysterone-controlled gene are differentially associated with ribosomes. *Gene.* 65:277–283.
4. Anderson, C. W., and E. Buzash-Pollert. 1985. Can ACG serve as an initiation codon for protein synthesis in eucaryotic cells? *Mol. Cell. Biol.* 5:3621–3624.
5. Arcari, P., R. Martinelli, and F. Salvatore. 1984. The complete sequence of a full length cDNA for human liver glyceraldehyde-3-phosphate dehydrogenase: evidence for multiple mRNA species. *Nucleic Acids Res.* 12:9179–9189.
6. Baim, S. B., and F. Sherman. 1988. mRNA structures influencing translation in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8:1591–1601.
7. Barbosa, M. S., and F. O. Wettstein. 1988. E2 of cottontail rabbit papillomavirus is a nuclear phosphoprotein translated from an mRNA encoding multiple open reading frames. *J. Virol.* 62:3242–3249.
8. Barkan, A. 1988. Proteins encoded by a complex chloropast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2637–2644.
9. Bayney, R. M., J. A. Rodkey, C. D. Bennett, A. Lu, and C. B. Pickett. 1987. Rat liver NADPH/quinone reductase: nucleotide sequence analysis of a quinone reductase cDNA clone and prediction of the amino acid sequence of the corresponding protein. *J. Biol. Chem.* 262:572–575.
10. Becerra, S. P., F. Koczot, P. Fabisch, and J. A. Rose. 1988. Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. *J. Virol.* 62:2745–2754.
11. Becerra, S. P., J. A. Rose, M. Hardy, B. M. Baroudy, and C. W. Anderson. 1985. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. *Proc. Natl. Acad. Sci. USA.* 82:7919–7923.
12. Benbrook, D., E. Lernhardt, and M. Pfahl. 1988. A new retinoic acid receptor identified from a hepatocellular carcinoma. *Nature (Lond.).* 333:669–672.
13. Bendena, W. G., M. Abouhaidar, and G. A. Mackie. 1985. Synthesis in vitro of the coat protein of papaya mosaic virus. *Virology.* 140:257–268.
14. Ben-Neriah, Y., and A. R. Bauskin. 1988. Leukocytes express a novel gene encoding a putative transmembrane protein-kinase devoid of an extracellular domain. *Nature (Lond.).* 333:672–676.
15. Bergmann, J. E., and H. F. Lodish. 1979. Translation of capped and uncapped vesicular stomatitis virus and reovirus mRNAs. *J. Biol. Chem.* 254:459–468.

16. Berkner, K. L., and P. A. Sharp. 1985. Effect of the tripartite leader on synthesis of a non-viral protein in an adenovirus 5 recombinant. *Nucleic Acids Res.* 13:841–857.
17. Bingham, P. M., T.-B. Chou, I. Mims, and Z. Zachar. 1988. On/off regulation of gene expression at the level of splicing. *Trends Genet.* 4:134–138.
18. Bonner, T. I., N. J. Buckley, A. C. Young, and M. R. Brann. 1987. Identification of a family of muscarinic acetylcholine receptor genes. *Science (Wash. DC).* 237:527–532.
19. Bonthron, D., and S. H. Orkin. 1988. The human von Willebrand factor gene: structure of the 5′ region. *Eur. J. Biochem.* 171:51–57.
20. Bossy, B., M. Ballivet, and P. Spierer. 1988. Conservation of neural nicotinic acetylcholine receptors from *Drosophila* to vertebrate central nervous systems. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:611–618.
21. Bourdon, M. A., A. Oldberg, M. Pierschbacher, and E. Ruoslahti. 1985. Molecular cloning and sequence analysis of a chondroitin sulfate proteoglycan cDNA. *Proc. Natl. Acad. Sci. USA.* 82:1321–1325.
22. Bourdon, M. A., M. Shiga, and E. Ruoslahti. 1986. Identification from cDNA of the precursor form of a chondroitin sulfate proteoglycan core protein. *J. Biol. Chem.* 261:12534–12537.
23. Brabant, M., L. McConlogue, T. van Daalen Wetters, and P. Coffino. 1988. Mouse ornithine decarboxylase gene: cloning, structure, and expression. *Proc. Natl. Acad. Sci. USA.* 85:2200–2204.
24. Breier, G., G. R. Dressler, and P. Gruss. 1988. Primary structure and developmental expression pattern of *Hox* 3.1, a member of the murine *Hox* 3 homeobox gene cluster. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1329–1336.
25. Brennan, C. A., A. J. Dombroski, and T. Platt. 1987. Transcription termination factor rho is an RNA-DNA helicase. *Cell.* 48:945–952.
26. Bzik, D. J., W-B. Li, T. Horii, and J. Inselburg. 1987. Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. *Proc. Natl. Acad. Sci. USA.* 84: 8360–8364.
27. Cattaneo, R., A. Schmid, D. Eschle, K. Baczko, V. ter Meulen, and M. A. Billeter. 1988. Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell.* 55:255–265.
28. Chavrier, P., M. Zerial, P. Lemaire, J. Almendral, R. Bravo, and P. Charnay. 1988. A gene encoding a protein with zinc fingers is activated during G_0/G_1 transition in cultured cells. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:29–35.
29. Cigan, A. M., and T. F. Donahue. 1987. Sequence and structural features associated with translational initiator regions in yeast—a review. *Gene.* 59:1–18.
30. Cigan, A. M., L. Feng, and T. F. Donahue. 1988. $tRNA_i^{met}$ functions in directing the scanning ribosome to the start site of translation. *Science (Wash. DC).* 242:93–97.
31. Cigan, A. M., E. K. Pabich, and T. F. Donahue. 1988. Mutational analysis of the *HIS4* translational initiator region in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8:2964–2975.
32. Clements, J. M., T. M. Laz, and F. Sherman. 1988. Efficiency of translation initiation by non-AUG codons in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8:4533–4536.
33. Cobrinik, D., L. Soskey, and J. Leis. 1988. A retroviral RNA secondary structure required for efficient initiation of reverse transcription. *J. Virol.* 62:3622–3630.
34. Cox, J. V., and E. Lazarides. 1988. Alternative primary structures in the trans-membrane domain of the chicken erythroid anion transporter. *Mol. Cell. Biol.* 8:1327–1335.
35. Craig, N., and M. Kostura. 1983. Inhibition of protein synthesis in CHO cells by actinomycin D: lesion occurs after 40S initiation complex formation. *Biochemistry.* 22:6064–6071.
36. Crum, C., J. D. Johnson, A. Nelson, and D. Roth. 1988. Complementary oligodeoxy-nucleotide mediated inhibition of tobacco mosaic virus RNA translation *in vitro*. *Nucleic Acids Res.* 16:4569–4581.
37. Curran, J., and D. Kolakofsky. 1988. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:245–251.
38. Curran, J., and D. Kolakofsky. 1988. Scanning independent ribosomal initiation of the Sendai virus X protein. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2869–2874.
39. Dabrowski, C., and J. C. Alwine. 1988. Translational control of synthesis of SV40 late proteins from polycistronic 19S late mRNA. *J. Virol.* 62:3182–3192.
40. DeGreve, J., J. Battey, J. Fedorko, M. Birrer, G. Evan, F. Kaye, E. Sausville, and J. Minna. 1988. The human L-*myc* gene encodes multiple nuclear phosphoproteins from alternatively processed mRNAs. *Mol. Cell. Biol.* 8:4381–4388.
41. Delli Bovi, P., A. M. Curatola, F. G. Kern, A. Greco, M. Ittmann, and C. Basilico. 1987. An oncogene isolated by transfection of Kaposi's sarcoma DNA encodes a growth factor that is a member of the FGF family. *Cell.* 50:729–737.
42. DePinho, R. A., K. S. Hatton, A. Tesfaye, G. D. Yancopoulos, and F. W. Alt. 1987. The human *myc* gene family: structure and activity of L-*myc* and an L-*myc* pseudo-gene. *Genes Dev.* 1:1311–1326.
43. Dixon, L. K., and T. Hohn. 1984. Initiation of translation of the cauliflower mosaic virus genome from a polycistronic mRNA: evidence from deletion mutagenesis. *EMBO (Eur. Mol. Biol. Organ.) J.* 3:2731–2736.
44. Donahue, T. F., and A. M. Cigan. 1988. Genetic selection for mutations that reduce or abolish ribosomal recognition of the *HIS4* translational initiator region. *Mol. Cell. Biol.* 8:2955–2963.
45. Donahue, T. F., A. M. Cigan, E. K. Pabich, and B. C. Valavicius. 1988. Mutations at a Zn(II) finger motif in the yeast eIF-2β gene alter ribosomal start-site selection during the scanning process. *Cell.* 54: 621–632.
46. Dorner, A. J., B. L. Semler, R. J. Jackson, R. Hanecak, E. Duprey, and E. Wimmer. 1984. In vitro translation of poliovirus RNA: utilization of internal initiation sites in reticulocyte lysate. *J. Virol.* 50:507–514.
47. Dorsky, D. I., and C. S. Crumpacker. 1988. Expression of herpes simplex virus type I DNA polymerase gene by in vitro translation and effects of gene deletions on activity. *J. Virol.* 62:3224–3232.
48. Driscoll, D. M., and J. G. Williams. 1987. Two divergently transcribed genes of *Dictyostelium discoideum* are cyclic AMP-inducible and coregulated during development. *Mol. Cell. Biol.* 7:4482–4489.
49. Duncan, R., and P. Dobos. 1986. The nucleotide sequence of infectious pancreatic necrosis virus (IPNV) dsRNA segment A reveals one large ORF encoding a precursor polyprotein. *Nucleic Acids Res.* 14:5934.
50. Duncan, R., E. Nagy, P. J. Krell, and P. Dobos. 1987. Synthesis of the infectious pancreatic necrosis virus polyprotein, detection of a virus-encoded protease, and fine structure mapping of genome segment A coding regions. *J. Virol.* 61:3655–3664.
51. Early, A. E., J. G. Williams, H. E. Meyer, S. B. Por, E. Smith, K. L. Williams, and A. A. Gooley. 1988. Structural characterization of *Dictyostelium discoideum* pre-spore-specific gene D19 and of its product, cell surface glycoprotein PsA. *Mol. Cell. Biol.* 8:3458–3466.
52. Edenberg, H. J., K. Zhang, K. Fong, W. F. Bosron, and T-K. Li. 1985. Cloning and sequencing of cDNA encoding the complete mouse liver alcohol dehydrogenase. *Proc. Natl. Acad. Sci. USA.* 82:2262–2266.
53. Ellis, S. B., M. E. Williams, N. R. Ways, R. Brenner, A. H. Sharp, A. T. Leung, K. P. Campbell, E. McKenna, W. J. Koch, A. Hui, A. Schwartz, and M. M. Harpold. 1988. Sequence and expression of mRNAs encoding the α_1 and α_2 subunits of a DHP-sensitive calcium channel. *Science (Wash. DC).* 241:1661–1664.
54. Endlich, B., and S. Linn. 1985. The DNA restriction endonuclease of *Escherichia coli* B. I. Studies of the DNA translocation and the ATPase activities. *J. Biol. Chem.* 260:5720–5728.
55. Engler, P., and Ursula Storb. 1987. High-frequency deletional rearrangement of immunoglobulin κ gene segments introduced into a pre-B-cell line. *Proc. Natl. Acad. Sci. USA.* 84:4949–4953.
56. Ercolani, L., B. Florence, M. Denaro, and M. Alexander. 1988. Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. *J. Biol. Chem.* 263:15335–15341.
57. Fambrough, D. M., K. Takeyasu, J. Lippincott-Schwarz, and N. R. Siegel. 1988. Structure of LEP100, a glycoprotein that shuttles between lysosomes and the plasma membrane, deduced from the nucleotide sequence of the encoding cDNA. *J. Cell Biol.* 106:61–67.
58. Finocchiaro, G., M. Ito, Y. Ikeda, and K. Tanaka. 1988. Molecular cloning and nucleotide sequence of cDNAs encoding the α-subunit of human electron transfer flavoprotein. *J. Biol. Chem.* 263:15773–15780.
59. Flach, J., G. Lindquester, S. Berish, K. Hickman, and R. Devlin. 1986. Analysis of tropomyosin cDNAs isolated from skeletal and smooth muscle mRNA. *Nucleic Acids Res.* 14:9193–9211.
60. Frank, J., A. Verschoor, T. Wagenknecht, M. Radermacher, and J-M. Carazo. 1988. A new non-crystallographic image-processing technique reveals the architecture of ribosomes. *Trends Biochem. Sci.* 13:123–127.
61. Fresno, M., and D. Vazquez. 1980. Inhibitory effects of cap analogues on globin mRNA and EMC RNA translation in a reticulocyte cell-free system. *Eur. J. Biochem.* 103:125–132.
62. Freund, J. N., and B. P. Jarry. 1987. The *rudimentary* gene of *Drosophila melanogaster* encodes four enzymic functions. *J. Mol. Biol.* 193:1–13.
63. Froussard, P., P. Chastagner, G. Somme, A. Abadie, W. Greene, J. Theze, and S. Longacre. 1988. p55 Il-2 receptor mRNA precursors in murine T lymphocyte nuclei. *J. Immunol.* 141:1358–1364.
64. Fukasawa, K. M., and S. Li. 1986. Nucleotide sequence of the putative regulatory region of mouse lactate dehydrogenase-A gene. *Biochem. J.* 235:435–439.
65. Fütterer, J., K. Gordon, J. M. Bonneville, H. Sanfacon, B. Pisan, J. Penswick, and T. Hohn. 1988. The leading sequence of caulimovirus large RNA can be folded into a large stem-loop structure. *Nucleic Acids Res.* 16:8377–8390.
66. Gallie, D. R., D. E. Sleat, J. W. Watts, P. C. Turner, and T. M. A. Wilson. 1987. A comparison of eukaryotic viral 5′-leader sequences as enhancers of mRNA expression *in vivo*. *Nucleic Acids Res.* 15:8693–8711.
67. Gallie, D. R., D. E. Sleat, J. W. Watts, P. C. Turner, and T. M. A. Wilson. 1987. The 5′-leader sequence of tobacco mosaic virus RNA enhances the expression of foreign gene transcripts *in vitro* and *in vivo*. *Nucleic Acids Res.* 15:3257–3273.
68. Garvin, A. M., S. Pawar, J. D. Marth, and R. M. Perlmutter. 1988. Struc-

ture of the murine *lck* gene and its rearrangement in a murine lymphoma cell line. *Mol. Cell. Biol.* 8:3058–3064.
69. Geballe, A. P., and E. S. Mocarski. 1988. Translational control of cytomegalovirus gene expression is mediated by upstream AUG codons. *J. Virol.* 62:3334–3340.
70. Good, P. J., R. C. Welch, A. Barkan, M. B. Somasekhar, and J. E. Mertz. 1988. Both VP2 and VP3 are synthesized from each of the alternatively spliced late 19S RNA species of simian virus 40. *J. Virol.* 62:944–953.
71. Gottesdiener, K. M., B. A. Karpinski, T. Lindsten, J. L. Strominger, N. H. Jones, C. B. Thompson, and J. M. Leiden. 1988. Isolation and structural characterization of the human 4F2 heavy-chain gene, an inducible gene involved in T-lymphocyte activation. *Mol. Cell. Biol.* 8:3809–3819.
72. Grass, D. S., and J. L. Manley. 1987. Selective translation initiation on bicistronic simian virus 40 late mRNA. *J. Virol.* 61:2331–2335.
73. Gronemeyer, H., B. Turcotte, C. Quirin-Stricker, M. T. Bocquel, M. E. Meyer, Z. Krozowski, J. M. Jeltsch, T. Lerouge, J. M. Garnier, and P. Chambon. 1987. The chicken progesterone receptor: sequence, expression and functional analysis. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3985–3994.
74. Gupta, K. C., and S. Patwardhan. 1988. ACG, the initiator codon for a Sendai virus protein. *J. Biol. Chem.* 263:8553–8556.
75. Haarr, L., H. S. Marsden, C. M. Preston, J. R. Smiley, W. C. Summers, and W. P. Summers. 1985. Utilization of internal AUG codons for initiation of protein synthesis directed by mRNAs from normal and mutant genes encoding herpes simplex virus-specified thymidine kinase. *J. Virol.* 56:512–519.
76. Hackett, P. B., R. B. Peterson, C. Hensel, F. Albericio, S. I. Gunderson, A. C. Palmenberg, and G. Barany. 1986. Synthesis *in vitro* of a seven amino acid peptide encoded in the leader RNA of Rous sarcoma virus. *J. Mol. Biol.* 190:45–57.
77. Haley, J., N. Whittle, P. Bennett, D. Kinchington, A. Ullrich, and M. Waterfield. 1987. The human EGF receptor gene: structure of the 110 kb locus and identification of sequences regulating its transcription. *Oncogene Res.* 1:375–396.
78. Hall, L., R. K. Craig, M. R. Edbrooke, and P. N. Campbell. 1982. Comparison of the nucleotide sequence of cloned human and guinea pig pre-α-lactalbumin cDNA with that of chick pre-lysozyme cDNA suggests evolution from a common ancestral gene. *Nucleic Acids Res.* 10:3503–3515.
79. Haniu, M., H. Yuan, S. Chen, T. Iyanagi, T. D. Lee, and J. E. Shively. 1988. Structure-function relationship of NADPH:quinone reductase: characterization of NH_2-terminal blocking group and essential tyrosine and lysine residues. *Biochemistry.* 27:6877–6883.
80. Hassin, D., R. Korn, and M. S. Horwitz. 1986. A major internal initiation site for the *in vitro* translation of the adenovirus DNA polymerase. *Virology.* 155:214–224.
81. Heidecker, G., and J. Messing. 1986. Structural analysis of plant genes. *Annu. Rev. Plant Physiol.* 37:439–466.
82. Henikoff, S. 1987. Multifunctional polypeptides for purine de novo synthesis. *Bioessays.* 6:8–13.
83. Henning, W. D., C. Upton, G. McFadden, R. Majumdar, and W. A. Bridger. 1988. Cloning and sequencing of the cytoplasmic precursor to the α subunit of rat liver mitochondrial succinyl-CoA synthetase. *Proc. Natl. Acad. Sci. USA.* 85:1432–1436.
84. Herman, R. C. 1986. Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein. *J. Virol.* 58:797–804.
85. Herman, R. C. 1987. Characterization of the internal intiation of translation on the vesicular stomatitis virus phosphoprotein mRNA. *Biochemistry.* 26:8346–8350.
86. Hinnebusch, A. G. 1988. Mechanisms of gene regulation in the general control of amino acid biosynthesis in *Saccharomyces cerevisiae*. *Microbiol. Rev.* 52:248–273.
87. Hughes, S., K. Mellstrom, E. Kosik, F. Tamanoi, and J. Brugge. 1984. Mutation of a termination codon affects *src* initiation. *Mol. Cell. Biol.* 4:1738–1746.
88. Jackson, R. J. 1988. Picornaviruses break the rules. *Nature (Lond.).* 334: 292–293.
89. Jacobs, K., C. Shoemaker, R. Rudersdorf, S. D. Neill, R. J. Kaufman, A. Mufson, J. Seehra, S. S. Jones, R. Hewick, E. Fritsch, M. Kawakita, T. Shimizu, and T. Miyake. 1985. Isolation and characterization of genomic and cDNA clones of human erythropoietin. *Nature (Lond.).* 313:806–810.
90. Jang, S. K., H-G. Krausslich, M. Nicklin, G. Duke, A. C. Palmenberg, and E. Wimmer. 1988. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* 62:2636–2643.
91. Jay, G., S. Nomura, C. W. Anderson, and G. Khoury. 1981. Identification of the SV40 agnogene product: a DNA binding protein. *Nature (Lond.).* 291:346–349.
92. Johansen, H., D. Schümperli, and M. Rosenberg. 1984. Affecting gene expression by altering the length and sequence of the 5′ leader. *Proc. Natl. Acad. Sci. USA.* 81:7698–7702.
93. Johnson, P. J., J. M. Kooter, and P. Borst. 1987. Inactivation of transcription by UV irradiation of T. brucei provides evidence for a multicistronic transcription unit including a VSG gene. *Cell.* 51:273–281.
94. Joseph, L. J., M. Le Beau, G. Jamieson, S. Acharya, T. B. Shows, J. D. Rowley, and V. P. Sukhatme. 1988. Molecular cloning, sequencing and mapping of *EGR2*, a human early growth response gene encoding a protein with zinc-binding finger structure. *Proc. Natl. Acad. Sci. USA.* 85:7164–7168.
95. Kagimoto, M., K. Kagimoto, E. R. Simpson, and M. R. Waterman. 1988. Transcription of the bovine adrenodoxin gene produces two species of mRNA of which only one is translated into adrenodoxin. *J. Biol. Chem.* 263:8925–8928.
96. Katamine, S., V. Notario, C. D. Rao, T. Miki, M. Cheah, S. R. Tronick, and K. C. Robbins. 1988. Primary structure of the human *fgr* proto-oncogene product $p55^{c\text{-}fgr}$. *Mol. Cell. Biol.* 8:259–266.
97. Katz, A., and C. Kahana. 1988. Isolation and characterization of the mouse ornithine decarboxylase gene. *J. Biol. Chem.* 263:7604–7609.
98. Keesey, J. K., Jr., R. Bigelis, and G. R. Fink. 1979. The product of the *his4* gene cluster in *Saccharomyces cerevisiae*: a trifunctional polypeptide. *J. Biol. Chem.* 254:7427–7433.
99. Kelley, D. E., C. Coleclough, and R. P. Perry. 1982. Functional significance and evolutionary development of the 5′-terminal regions of immunoglobulin variable-region genes. *Cell.* 29:681–689.
100. Kemper, B., and L. Stolarsky. 1977. Dependence on potassium concentration of the inhibition of translation of mRNA by 7-methylguanosine 5′-phosphate. *Biochemistry.* 16:5676–5680.
101. Khalili, K., J. Brady, and G. Khoury. 1987. Translational regulation of SV40 early mRNA defines a new viral protein. *Cell.* 48:639–645.
102. Khalili, K., G. Khoury, and J. Brady. 1986. Spacing between simian virus 40 early transcriptional control sequences is important for regulation of early RNA synthesis and gene expression. *J. Virol.* 60:935–942.
103. Kim, H-R. C., N. S. Yew, W. Ansorge, H. Voss, C. Schwager, B. Vennstrom, M. Zenke, and J. D. Engel. 1988. Two different mRNAs are transcribed from a single genomic locus encoding the chicken erythrocyte anion transport proteins (band 3). *Mol. Cell. Biol.* 8:4416–4424.
104. King, A., and D. W. Melton. 1987. Characterisation of cDNA clones for hypoxanthine-guanine phosphoribosyltransferase from the human malarial parasite, *Plasmodium falciparum*: comparisons to the mammalian gene and protein. *Nucleic Acids Res.* 15:10469–10481.
105. King, A. M. Q., E. J. Stott, S. J. Langer, K. Young, L. A. Ball, and G. W. Wertz. 1987. Recombinant vaccinia viruses carrying the N gene of human respiratory syncytial virus: studies of gene expression in cell culture and immune response in mice. *J. Virol.* 61:2885–2890.
106. Klemenz, R., D. Hultmark, and W. J. Gehring. 1985. Selective translation of heat shock mRNA in *Drosophila melanogaster* depends on sequence information in the leader. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:2053–2060.
107. Kobilka, B. K., C. MacGregor, K. Daniel, T. S. Kobilka, M. G. Caron, and R. J. Lefkowitz. 1987. Functional activity and regulation of human β_2-adrenergic receptors expressed in *Xenopus* oocytes. *J. Biol. Chem.* 262:15796–15802.
108. Koide, T., D. Foster, S. Yoshitake, and E. W. Davie. 1986. Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA. *Biochemistry.* 25:2220–2225.
109. Konarska, M., W. Filipowicz, H. Domdey, and H. J. Gross. 1981. Binding of ribosomes to linear and circular forms of the 5′-terminal leader fragment of tobacco mosaic virus RNA. *Eur. J. Biochem.* 114:221–227.
110. Korneluk, R. G., D. J. Mahuran, K. Neote, M. H. Klavins, B. F. O'Dowd, M. Tropak, H. F. Willard, M-J. Anderson, J. A. Lowden, and R. A. Gravel. 1986. Isolation of cDNA clones coding for the α-subunit of human β-hexosaminidase. *J. Biol. Chem.* 261:8407–8413.
111. Kotani, H., M. Hoshimaru, H. Nawa, and S. Nakanishi. 1986. Structure and gene organization of bovine neuromedin K precursor. *Proc. Natl. Acad. Sci. USA.* 83:7074–7078.
112. Kozak, M. 1978. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell.* 15:1109–1123.
113. Kozak, M. 1979. Inability of circular mRNA to attach to eukaryotic ribosomes. *Nature (Lond.).* 280:82–85.
114. Kozak, M. 1979. Migration of 40S ribosomal subunits on messenger RNA when initiation is perturbed by lowering magnesium or adding drugs. *J. Biol. Chem.* 254:4731–4738.
115. Kozak, M. 1980. Role of ATP in binding and migration of 40S ribosomal subunits. *Cell.* 22:459–467.
116. Kozak, M. 1980. Binding of wheat germ ribosomes to fragmented viral mRNA. *J. Virol.* 35:748–756.
117. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9:5233–5252.
118. Kozak, M. 1983. Translation of insulin-related polypeptides from messenger RNAs with tandemly reiterated copies of the ribosome binding site. *Cell.* 34:971–978.
119. Kozak, M. 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes and organelles. *Microbiol. Rev.* 47:1–45.
120. Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* 12: 857–872.

121. Kozak, M. 1984. Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin. *Nucleic Acids Res.* 12:3873–3893.
122. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44:283–292.
123. Kozak, M. 1986. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA.* 83:2850–2854.
124. Kozak, M. 1986. Bifunctional messenger RNAs in eukaryotes. *Cell.* 47: 481–483.
125. Kozak, M. 1986. Regulation of protein synthesis in virus-infected animal cells. *Advances Virus Res.* 31:229–292.
126. Kozak, M. 1987. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* 196:947–950.
127. Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15:8125–8148.
128. Kozak, M. 1987. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol. Cell. Biol.* 7:3438–3445.
129. Kozak, M. 1988. Leader length and secondary structure modulate mRNA function under conditions of stress. *Mol. Cell. Biol.* 8:2737–2744.
130. Kozak, M. 1988. A profusion of controls. *J. Cell Biol.* 107:1–7.
131. Kozak, M., and A. J. Shatkin. 1978. Migration of 40S ribosomal subunits on messenger RNA in the presence of edeine. *J. Biol. Chem.* 253: 6568–6577.
132. Krieg, P., R. Strachan, E. Wallis, L. Tabe, and A. Colman. 1984. Efficient expression of cloned complementary DNAs for secretory proteins after injection into *Xenopus* oocytes. *J. Mol. Biol.* 180:615–643.
133. Kubo, T., K. Fukuda, A. Mikami, A. Maeda, H. Takahashi, M. Mishina, T. Haga, K. Haga, A. Ichiyama, K. Kangawa, M. Kojima, H. Matsuo, T. Hirose, and S. Numa. 1986. Cloning, sequencing and expression of complementary DNA encoding the muscarinic acetylcholine receptor. *Nature (Lond.).* 323:411–416.
134. Kurokawa, T., M. Seno, and K. Igarashi. 1988. Nucleotide sequence of rat basic fibroblast growth factor cDNA. *Nucleic Acids Res.* 16:5201.
135. Kusumoto, H., S. Hirosawa, J. P. Salier, F. S. Hagen, and K. Kurachi. 1988. Human genes for complement components Clr and Cls in a close tail-to-tail arrangement. *Proc. Natl. Acad. Sci. USA.* 85:7307–7311.
136. Kwiatkowski, D. J., and G. A. P. Bruns. 1988. Human profilin: molecular cloning, sequence comparison, and chromosomal analysis. *J. Biol. Chem.* 263:5910–5915.
137. Kwoh, T. J., D. Y. Kwoh, A. W. McCue, G. R. Davis, D. Patrick, and T. R. Gingeras. 1986. Introduction and expression of the bacterial *Pae*R7 methylase gene in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 83:7713–7717.
138. Ladner, M. B., G. A. Martin, J. A. Noble, D. M. Nikoloff, R. Tal, E. S. Kawasaki, and T. J. White. 1987. Human CSF-1: gene structure and alternative splicing of mRNA precursors. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:2693–2698.
139. Laird, J. E., L. Jack, L. Hall, A. P. Boulton, D. Parker, and R. K. Craig. 1988. Structure and expression of the guinea pig α-lactalbumin gene. *Biochem. J.* 254:85–94.
140. Lawrence, C. B. 1980. Activation of an internal initiation site for protein synthesis during in vitro translation. *Nucleic Acids Res.* 8:1307–1317.
141. Lee, W-H., R. Bookstein, F. Hong, L-J. Young, J-Y. Shew, E. Y-H. P. Lee. 1987. Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science (Wash. DC).* 235:1394–1399.
142. Lee, W-H., J-Y. Shew, F. D. Hong, T. W. Sery, L. A. Donoso, L-J. Young, R. Bookstein, and E. Y-H. P. Lee. 1987. The retinoblastoma susceptibility gene encodes a nuclear phosphoprotein associated with DNA binding activity. *Nature (Lond.).* 329:642–645.
143. Lennon, G. G., and R. P. Perry. 1985. C_μ-containing transcripts initiate heterogeneously within the *IgH* enhancer region and contain a novel 5'-nontranslatable exon. *Nature (Lond.).* 318:475–478.
144. Lentz, M. R., and G. M. Air. 1986. Loss of enzyme activity in a site-directed mutant of influenza neuraminidase compared to expressed wild-type protein. *Virology.* 148:74–83.
145. Leung, D. W., S. A. Spencer, G. Cachianes, R. G. Hammonds, C. Collins, W. J. Henzel, R. Barnard, M. J. Waters, and W. I. Wood. 1987. Growth hormone receptor and serum binding protein: purification, cloning and expression. *Nature (Lond.).* 330:537–543.
146. Levitan, E. S., P. R. Schofield, D. R. Burt, L. M. Rhee, W. Wisden, M. Kohler, N. Fujita, H. F. Rodriguez, A. Stephenson, M. G. Darlison, E. A. Barnard, and P. H. Seeburg. 1988. Structural and functional basis for $GABA_A$ receptor heterogeneity. *Nature (Lond.).* 335:76–79.
147. Li, S., H. F. Tiano, K. M. Fukasawa, K. Yagi, M. Shimizu, F. S. Sharief, Y. Nakashima, and Y. E. Pan. 1985. Protein structure and gene organization of mouse lactate dehydrogenase-A isozyme. *Eur. J. Biochem.* 149:215–225.
148. Liu, C-C., C. C. Simonsen, and A. D. Levinson. 1984. Initiation of translation at internal AUG codons in mammalian cells. *Nature (Lond.).* 309: 82–85.
149. Liu, F-T., K. Albrandt, and M. W. Robertson. 1988. cDNA heterogeneity suggests structural variants related to the high-affinity IgE receptor. *Proc. Natl. Acad. Sci. USA.* 85:5639–5643.
150. Logan, J., and T. Shenk. 1984. Adenovirus tripartite leader sequence enhances translation of mRNAs late after infection. *Proc. Natl. Acad. Sci. USA.* 81:3655–3659.
151. Lumadue, J. A., A. B. Glick, and F. H. Ruddle. 1987. Cloning, sequence analysis, and expression of the large subunit of the human lymphocyte activation antigen 4F2. *Proc. Natl. Acad. Sci. USA.* 84:9204–9208.
152. Mackinnon, C. M., P. E. Carter, S. J. Smyth, B. Dunbar, and J. E. Fothergill. 1987. Molecular cloning of cDNA for human complement component Cls. *Eur. J. Biochem.* 169:547–553.
153. Mahuran, D. J., K. Neote, M. H. Klavins, A. Leung, and R. A. Gravel. 1988. Proteolytic processing of pro-α and pro-β precursors from human β-hexosaminidase. *J. Biol. Chem.* 263:4612–4618.
154. Mangin, M., A. C. Webb, B. E. Dreyer, J. T. Posillico, K. Ikeda, E. C. Weir, A. Stewart, N. Bander, L. Milstone, D. Barton, U. Francke, and A. E. Broadus. 1988. Identification of a cDNA encoding a parathyroid hormone-like peptide from a human tumor associated with humoral hypercalcemia of malignancy. *Proc. Natl. Acad. Sci. USA.* 85:597–601.
155. Mansour, S. L., and G. R. Martin. 1988. Four classes of mRNA are expressed from the mouse *int*-2 gene, a member of the FGF gene family. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2035–2041.
156. Marth, J. D., R. W. Overell, K. E. Meier, E. G. Krebs, and R. M. Perlmutter. 1988. Translational activation of the *lck* proto-oncogene. *Nature (Lond.).* 332:171–173.
157. Martini, G., D. Toniolo, T. Vulliamy, L. Luzzatto, R. Dono, G. Viglietto, G. Paonessa, M. D'Urso, and M. G. Persico. 1986. Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:1849–1855.
158. Mattick, J. S., Z. E. Zehner, M. A. Calabro, and S. J. Wakil. 1981. The isolation and characterization of fatty-acid synthetase mRNA from rat mammary gland. *Eur. J. Biochem.* 114:643–651.
159. McClelland, A., L. C. Kühn, and F. H. Ruddle. 1984. The human transferrin receptor gene: genomic organization, and the complete primary structure of the receptor deduced from a cDNA sequence. *Cell.* 39: 267–274.
160. McGarry, T. J., and S. Lindquist. 1985. The preferential translation of Drosophila hsp70 mRNA requires sequences in the untranslated leader. *Cell.* 42:903–911.
161. McIvor, R. S., M. J. Johnson, A. D. Miller, S. Pitts, S. R. Williams, D. Valerio, D. W. Martin, Jr., and I. M. Verma. 1987. Human purine nucleoside phosphorylase and adenosine deaminase: gene transfer into cultured cells and murine hematopoietic stem cells by using recombinant amphotropic retroviruses. *Mol. Cell. Biol.* 7:838–846.
162. Mertens, P. P. C., and P. Dobos. 1982. Messenger RNA of infectious pancreatic necrosis virus is polycistronic. *Nature (Lond.).* 297:243–246.
163. Misrahi, M., H. Loosfelt, M. Atger, C. Mériel, V. Zerah, P. Dessen, and E. Milgrom. 1988. Organization of the entire rabbit progesterone receptor mRNA and of the promoter and 5' flanking region of the gene. *Nucleic Acids Res.* 16:5459–5472.
164. Mitchell, D. B., and G. W. Both. 1988. Simian rotavirus SA11 segment 11 contains overlapping reading frames. *Nucleic Acids Res.* 16:6244.
165. Morch, M.-D., J.-C. Boyer, and A.-L. Haenni. 1988. Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. *Nucleic Acids Res.* 16:6157–6173.
166. Morgan, D. O., J. C. Edman, D. N. Standring, V. A. Fried, M. C. Smith, R. A. Roth, and W. J. Rutter. 1987. Insulin-like growth factor II receptor as a multifunctional binding protein. *Nature (Lond.).* 329:301–307.
167. Morlé, F., B. Lopez, T. Henni, and J. Godet. 1985. α-Thalassaemia associated with the deletion of two nucleotides at position −2 and −3 preceding the AUG codon. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1245–1250.
168. Mueller, P. P., and A. G. Hinnebusch. 1986. Multiple upstream AUG codons mediate translational control of *GCN4*. *Cell.* 45:201–207.
169. Murphy, A. J. M., and A. Efstratiadis. 1987. Cloning vectors for expression of cDNA libraries in mammalian cells. *Proc. Natl. Acad. Sci. USA.* 84:8277–8281.
170. Murphy, S. P., J. Garbern, W. F. Odenwald, R. A. Lazzarini, and E. Linney. 1988. Differential expression of the homeobox gene *Hox-1.3* in F9 embryonal carcinoma cells. *Proc. Natl. Acad. Sci. USA.* 85:5587–5591.
171. Murray, M. B., N. D. Zilz, N. L. McCreary, M. J. MacDonald, and H. C. Towle. 1988. Isolation and characterization of rat cDNA clones for two distinct thyroid hormone receptors. *J. Biol. Chem.* 263:12770–12777.
172. Mutzel, R., M.-L. Lacombe, M.-N. Simon, J. deGunzburg, and M. Veron. 1987. Cloning and cDNA sequence of the regulatory subunit of cAMP-dependent protein kinase from *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA.* 84:6–10.
173. Nagashima, K., M. Yoshida, and M. Seiki. 1986. A single species of pX mRNA of human T-cell leukemia virus type I encodes trans-activator $p40^x$ and two other phosphoproteins. *J. Virol.* 60:394–399.
174. Namen, A. E., S. Lupton, K. Hjerrild, J. Wignall, D. Mochizuki, A. Schmierer, B. Mosley, C. March, D. Urdal, S. Gillis, D. Cosman, and R. G. Goodwin. 1988. Stimulation of B-cell progenitors by cloned murine interleukin-7. *Nature (Lond.).* 333:571–573.
175. Nielsen, P. J., and H. Trachsel. 1988. The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which

are differentially expressed. *EMBO (Eur. Mol. Biol. Organ.) J.* 7: 2097–2105.
176. Nielsen, P. J., G. K. McMaster, and H. Trachsel. 1985. Cloning of eukaryotic protein synthesis initiation factor genes: isolation and characterization of cDNA clones for encoding factor eIF-4A. *Nucleic Acids Res.* 13:6867–6880.
177. Noda, M., T. Ikeda, T. Kayano, H. Suzuki, H. Takeshima, M. Kurasaki, H. Takahashi, and S. Numa. 1986. Existence of distinct sodium channel messenger RNAs in rat brain. *Nature (Lond.).* 320:188–192.
178. Ohno, S., Y. Akita, Y. Konno, S. Imajoh, and K. Suzuki. 1988. A novel phorbol ester receptor/protein kinase, nPKC, distantly related to the protein kinase C family. *Cell.* 53:731–741.
179. Ohta, S., H. Tomura, K. Matsuda, and Y. Kagawa. 1988. Gene structure of the human mitochondrial adenosine triphosphate synthase β subunit. *J. Biol. Chem.* 263:11257–11262.
180. Olate, J., R. Mattera, J. Codina, and L. Birnbaumer. 1988. Reticulocyte lysates synthesize an active α subunit of the stimulatory G protein G_s. *J. Biol. Chem.* 263:10394–10400.
181. Oprian, D. D., R. S. Molday, R. J. Kaufman, and H. G. Khorana. 1987. Expression of a synthetic bovine rhodopsin gene in monkey kidney cells. *Proc. Natl. Acad. Sci. USA.* 84:8874–8878.
182. O'Rear, J. J., and H. M. Temin. 1982. Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, and avian retrovirus. *Proc. Natl. Acad. Sci. USA.* 79:1230–1234.
183. Orkin, S. H. 1987. The pitfalls of sequencing: revised sequence of the X-CGD gene. *Trends. Genet.* 3:207.
184. Osborne, W. R. A., and A. D. Miller. 1988. Design of vectors for efficient expression of human purine nucleoside phosphorylase in skin fibroblasts from enzyme-deficient humans. *Proc. Natl. Acad. Sci. USA.* 85:6851–6855.
185. Ozawa, K., J. Ayub, and N. Young. 1988. Translational regulation of B19 parvovirus capsid protein production by multiple upstream AUG triplets. *J. Biol. Chem.* 263:10922–10926.
186. Ozawa, M., R.-P. Huang, T. Furukawa, and T. Muramatsu. 1988. A teratocarcinoma glycoprotein carrying a developmentally regulated carbohydrate marker is a member of the immunoglobulin gene superfamily. *J. Biol. Chem.* 263:3059–3062.
187. Pachnis, V., C. I. Brannan, and S. M. Tilghman. 1988. The structure and expression of a novel gene activated in early mouse embryogenesis. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:673–681.
188. Paluh, J. L., M. J. Orbach, T. L. Legerton, and C. Yanofsky. 1988. The cross-pathway control gene of *Neurospora crassa, cpc-1*, encodes a protein similar to GCN4 of yeast and the DNA-binding domain of the oncogene v-*jun*-encoded protein. *Proc. Natl. Acad. Sci. USA.* 85:3728–3732.
189. Park, M., M. Dean, K. Kaul, M. J. Braun, M. A. Gonda, and G. Vande Woude. 1987. Sequence of *MET* protooncogene cDNA has features characteristic of the tyrosine kinase family of growth-factor receptors. *Proc. Natl. Acad. Sci. USA.* 84:6379–6383.
190. Patel, D., and D. J. Pickup. 1987. Messenger RNAs of a strongly-expressed late gene of cowpox virus contain 5′-terminal poly(A) sequences. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:3787–3794.
191. Pathak, V. K., P. J. Nielsen, H. Trachsel, and J. W. B. Hershey. 1988. Structure of the β subunit of translational initiation factor eIF2. *Cell.* 54:633–639.
192. Peabody, D. S. 1987. Translation initiation at an ACG triplet in mammalian cells. *J. Biol. Chem.* 262:11847–11851.
193. Peabody, D. S., S. Subramani, and P. Berg. 1986. Effect of upstream reading frames on translation efficiency in SV40 recombinants. *Mol. Cell. Biol.* 6:2704–2711.
194. Pearson-White, S. H., and C. P. Emerson, Jr. 1987. A novel hybrid α-tropomyosin in fibroblasts is produced by alternative splicing of transcripts from the skeletal muscle α-tropomyosin gene. *J. Biol. Chem.* 262:15998–16010.
195. Pelham, H. R. B. 1979. Translation of fragmented viral RNA in vitro. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 100:195–199.
196. Pelletier, J., and N. Sonenberg. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature (Lond.).* 334:320–325.
197. Peralta, E. G., J. W. Winslow, G. L. Peterson, D. H. Smith, A. Ashkenazi, J. Ramachandran, M. I. Schimerlik, and D. J. Capon. 1987. Primary structure and biochemical properties of an M_2 muscarinic receptor. *Science (Wash. DC).* 236:600–605.
198. Perez, L., J. W. Wills, and E. Hunter. 1987. Expression of the Rous sarcoma virus *env* gene from a simian virus 40 late-region replacement vector: effects of upstream initiation codons. *J. Virol.* 61:1276–1281.
199. Peterson, C. A., and J. Piatigorsky. 1986. Preferential conservation of the globular domains of the βA3/A1-crystallin polypeptide of the chicken eye lens. *Gene.* 45:139–147.
200. Phillips, B. A., and A. Emmert. 1986. Modulation of the expression of poliovirus proteins in reticulocyte lysates. *Virology.* 148:255–267.
201. Pinck, L., A. Franck, and C. Fritsch. 1979. Formation of ribosome-RNA initiation complexes with alfalfa mosaic virus RNA 4 and RNA 3. *Nucleic Acids Res.* 7:151–166.
202. Pinck, M., C. Fritsch, M. Ravelonandro, C. Thivent, and L. Pinck. 1981. Binding of ribosomes to the 5′ leader sequence (N=258) of RNA 3 from alfalfa mosaic virus. *Nucleic Acids Res.* 9:1087–1100.
203. Pirastu, M., G. Saglio, J. C. Chang, A. Cao, and Y. Wai Kan. 1984. Initiation codon mutation as a cause of α thalassemia. *J. Biol. Chem.* 259: 12315–12317.
204. Propst, F., M. P. Rosenberg, A. Iyer, K. Kaul, and G. F. Vande Woude. 1987. c-*mos* proto-oncogene RNA transcripts in mouse tissues: structural features, developmental regulation, and localization in specific cell types. *Mol. Cell. Biol.* 7:1629–1637.
205. Pytela, R. 1988. Amino acid sequence of the murine Mac-1 α chain reveals homology with the integrin family and an additional domain related to von Willebrand factor. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1371–1378.
206. Quintrell, N., R. Lebo, H. Varmus, J. M. Bishop, M. J. Pettenati, M. M. LeBeau, M. O. Diaz, and J. D. Rowley. 1987. Identification of a human gene (*HCK*) that encodes a protein-tyrosine kinase and is expressed in hemopoietic cells. *Mol. Cell. Biol.* 7:2267–2275.
207. Racaniello, V. R., and D. Baltimore. 1981. Molecular cloning of poliovirus cDNA and determination of the complete nucleotide sequence of the viral genome. *Proc. Natl. Acad. Sci. USA.* 78:4887–4891.
208. Ragheb, J. A., and R. P. Dottin. 1987. Structure and sequence of a UDP glucose pyrophosphorylase gene of *Dictyostelium discoideum. Nucleic Acids Res.* 15:3891–3906.
209. Rao, C. D., M. Pech, K. C. Robbins, and S. A. Aaronson. 1988. The 5′ untranslated sequence of the c-*sis*/platelet-derived growth factor 2 transcript is a potent translational inhibitor. *Mol. Cell. Biol.* 8:284–292.
210. Reisman, D., M. Greenberg, and V. Rotter. 1988. Human p53 oncogene contains one promoter upstream of exon 1 and a second, stronger promoter within intron 1. *Proc. Natl. Acad. Sci. USA.* 85:5146–5150.
211. Resnick, J., and T. Shenk. 1986. Simian virus 40 agnoprotein facilitates normal nuclear location of the major capsid polypeptide and cell-to-cell spread of virus. *J. Virol.* 60:1098–1106.
212. Reynolds, G. A., J. L. Goldstein, and M. S. Brown. 1985. Multiple mRNAs for 3-hydroxy-3-methylglutaryl coenzyme A reductase determined by multiple transcription initiation sites and intron splicing sites in the 5′-untranslated region. *J. Biol. Chem.* 260:10369–10377.
213. Rice, C. M., E. G. Struss, and J. H. Strauss. 1986. Structure of the flavivirus genome. *In* The Togaviridae and Flaviviridae S. Schlesinger and M. J. Schlesinger, editors. Plenum Publishing Corp., New York. 279–326.
214. Rogers, S. G., R. T. Fraley, R. B. Horsch, A. D. Levine, J. S. Flick, L. A. Brand, C. L. Fink, T. Mozer, K. O'Connell, and P. R. Sanders. 1985. Evidence for ribosome scanning during translation initiation of mRNAs in transformed plant cells. *Plant Mol. Biol. Reporter.* 3:111–116.
215. Ron, D., S. R. Tronick, S. A. Aaronson, and A. Eva. 1988. Molecular cloning and characterization of the human *dbl* proto-oncogene: evidence that its overexpression is sufficient to transform NIH/3T3 cells. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:2465–2473.
216. Rorsman, F., M. Bywater, T. J. Knott, J. Scott, and C. Betsholtz. 1988. Structural characterization of the human platelet-derived growth factor A-chain cDNA and gene: alternative exon usage predicts two different precursor proteins. *Mol. Cell. Biol.* 8:571–577.
217. Rosa, F., H. Berissi, J. Weissenbach, L. Maroteaux, M. Fellous, and M. Revel. 1983. The β_2-microglobulin mRNA in human Daudi cells has a mutated initiation codon but is still inducible by interferon. *EMBO (Eur. Mol. Biol. Organ.) J.* 2:239–243.
218. Rosel, J. L., P. L. Earl, J. P. Weir, and B. Moss. 1986. Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the *Hin*dIII H genome fragment. *J. Virol.* 60:436–449.
219. Rouault, T. A., M. W. Hentze, S. W. Caughman, J. B. Harford, and R. D. Klausner. 1988. Binding of a cytosolic protein to the iron-responsive element of human ferritin messenger RNA. *Science (Wash. DC).* 241:1207–1213.
220. Roussou, I., G. Thireos, and B. M. Hauge. 1988. Transcriptional-translational regulatory circuit in *Saccharomyces cerevisiae* which involves the *GCN4* transcriptional activator and the GCN2 protein kinase. *Mol. Cell. Biol.* 8:2132–2139.
221. Sangar, D. V., S. E. Newton, D. J. Rowlands, and B. E. Clarke. 1987. All foot-and-mouth disease virus serotypes initiate protein synthesis at two separate AUGs. *Nucleic Acids Res.* 15:3305–3315.
222. Sarver, N., G. A. Ricca, J. Link, M. H. Nathan, J. Newman, and W. N. Drohan. 1987. Stable expression of recombinant factor VIII molecules using a bovine papilloma-virus vector. *DNA.* 6:553–564.
223. Schimenti, J., J. A. Cebra-Thomas, C. L. Decker, S. D. Islam, S. H. Pilder, and L. M. Silver. 1988. A candidate gene family for the mouse *t complex responder (Tcr)* locus responsible for haploid effects on sperm function. *Cell.* 55:71–78.
224. Schneider, C., M. J. Owen, D. Banville, and J. G. Williams. 1984. Primary structure of human transferrin receptor deduced from the mRNA sequence. *Nature (Lond.).* 311:675–678.
225. Schulz, P., R. Stucka, H. Feldman, G. Combriato, H.-G. Klobeck, and F. Fittler. 1988. Sequence of a cDNA clone encompassing the complete mature human prostate specific antigen (PSA) and an unspliced leader

sequence. *Nucleic Acids Res.* 16:6226.
226. Schwer, B., P. Visca, J. C. Vos, and H. G. Stunnenberg. 1987. Discontinuous transcription or RNA processing of vaccinia virus late messengers results in a 5′ poly(A) leader. *Cell.* 50:163–169.
227. Seal, S. N., A. Schmidt, M. Tomaszewski, and A. Marcus. 1978. Inhibition of noncapped mRNA translation by the cap analogue, 7-methylguanosine-5′-phosphate. *Biochem. Biophys. Res. Commun.* 82:553–559.
228. Sedman, S. A., and J. E. Mertz. 1988. Mechanisms of synthesis of virion proteins from the functionally bigenic late mRNAs of simian virus 40. *J. Virol.* 62:954–961.
229. Sehgal, A., N. Patil, and M. Chao. 1988. A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell. Biol.* 8:3160–3167.
230. Seto, M., U. Jaeger, R. D. Hockett, W. Graninger, S. Bennett, P. Goldman, and S. J. Korsmeyer. 1988. Alternative promoters and exons, somatic mutation and deregulation of the *Bcl-2-Ig* fusion gene in lymphoma. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:123–131.
231. Shaper, N. L., G. F. Hollis, J. G. Douglas, I. R. Kirsch, and J. H. Shaper. 1988. Characterization of the full-length cDNA for murine β-1, 4-galactosyltransferase. *J. Biol. Chem.* 263:10420–10428.
232. Shatkin, A. J. 1976. Capping of eucaryotic mRNAs. *Cell.* 9:645–653.
233. Shatkin, A. J. 1985. mRNA cap binding proteins: essential factors for initiating translation. *Cell.* 40:223–224.
234. Shaw, M. W., P. W. Choppin, and R. A. Lamb. 1983. A previously unrecognized influenza B virus glycoprotein from a bicistronic mRNA that also encodes the viral neuraminidase. *Proc. Natl. Acad. Sci. USA.* 80: 4879–4883.
235. Sherman, F., J. W. Stewart, and A. M. Schweingruber. 1980. Mutants of yeast initiating translation of iso-1-cytochrome c within a region spanning 37 nucleotides. *Cell.* 20:215–222.
236. Shimatsu, A., and P. Rotwein. 1987. Sequence of two rat insulin-like growth factor I mRNAs differing within the 5′ untranslated region. *Nucleic Acids Res.* 15:7196.
237. Shu, L., M. S. Horwitz, and J. A. Engler. 1987. Expression of enzymatically active adenovirus DNA polymerase from cloned DNA requires sequences upstream of the main open reading frame. *Virology.* 161:520–526.
238. Sillekens, P. T. G., R. P. Beijer, W. J. Habets, and W. J. van Venrooij. 1988. Human U1 snRNP-specific C protein: complete cDNA and protein sequence and identification of a multigene family in mammals. *Nucleic Acids Res.* 16:8307–8321.
239. Simeone, A., F. Mavilio, D. Acampora, A. Giampaolo, A. Faiella, V. Zappavigna, M. D'Esposito, M. Pannese, G. Russo, E. Boncinelli, and C. Peschle. 1987. Two human homeobox genes, *c1* and *c8*: structure analysis and expression in embryonic development. *Proc. Natl. Acad. Sci. USA.* 84:4914–4918.
240. Simonsen, C. C., M. Walter, and A. D. Levinson. 1988. Expression of the plasmid-encoded type I dihydrofolate reductase gene in cultured mammalian cells: a novel selectable marker. *Nucleic Acids Res.* 16: 2235–2246.
241. Siomi, H., H. Shida, S. H. Nam, T. Nosaka, M. Maki, and M. Hatanaka. 1988. Sequence requirements for nucleolar localization of human T cell leukemia virus type I pX protein, which regulates viral RNA processing. *Cell.* 55:197–209.
242. Sleat, D. E., R. Hull, P. C. Turner, and T. M. A. Wilson. 1988. Studies on the mechanism of translational enhancement by the 5′-leader sequence of tobacco mosaic virus RNA. *Eur. J. Biochem.* 175:75–86.
243. Smith, A. E. 1977. Cryptic initiation sites in eukaryotic virus mRNAs. *Fed. Eur. Biol. Soc. Symp.* 43:37–46.
244. Smith, G. E., M. D. Summers, and M. J. Fraser. 1983. Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Mol. Cell. Biol.* 3:2156–2165.
245. Smythe, J. A., R. L. Coppel, G. V. Brown, R. Ramasamy, D. J. Kemp, and R. F. Anders. 1988. Identification of two integral membrane proteins of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA.* 85:5195–5199.
246. Sonenberg, N., and A. J. Shatkin. 1978. Nonspecific effect of m7GMP on protein-RNA interactions. *J. Biol. Chem.* 253:6630–6632.
247. Spritz, R. A., K. Strunk, C. S. Surowy, S. O. Hoch, D. E. Barton, and U. Francke. 1987. The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding. *Nucleic Acids Res.* 15:10373–10391.
248. Stirzaker, S. C., P. L. Whitfeld, D. L. Christie, A. R. Bellamy, and G. W. Both. 1987. Processing of rotavirus glycoprotein VP7: implications for the retention of the protein in the endoplasmic reticulum. *J. Cell Biol.* 105:2897–2903.
249. Stolle, C. A., and E. J. Benz, Jr. 1988. Cellular factor affecting the stability of β-globin mRNA. *Gene.* 62:65–74.
250. Strick, C. A., and T. D. Fox. 1987. *Saccharomyces cerevisiae* positive regulatory gene *PET111* encodes a mitochondrial protein that is translated from an mRNA with a long 5′ leader. *Mol. Cell. Biol.* 7:2728–2734.
251. Strubin, M., E. O. Long, and B. Mach. 1986. Two forms of the Ia antigen-associated invariant chain result from alternative initiations at two in-phase AUGs. *Cell.* 47:619–625.
252. Sukhatme, V. P., X. Cao, L. Chang, C.-H. Tsai-Morris, D. Stamenkovich, P. Ferreira, D. Cohen, S. Edwards, T. Shows, T. Curran, M. Le Beau, and E. D. Adamson. 1988. A zinc finger-encoding gene coregulated with *c-fos* during growth and differentiation, and after cellular depolarization. *Cell.* 53:37–43.
253. Takebe, Y., M. Seiki, J.-I. Fujisawa, P. Hoy, K. Yokota, K.-I. Arai, M. Yoshida, and N. Arai. 1988. SRα promoter: an efficient and versatile mammalian cDNA expression system composed of the SV40 early promoter and the R-U5 segment of HTLV type I long terminal repeat. *Mol. Cell. Biol.* 8:466–472.
254. Takizawa, T., I.-Y. Huang, T. Ikuta, and A. Yoshida. 1986. Human glucose-6-phosphate dehydrogenase: primary structure and cDNA cloning. *Proc. Natl. Acad. Sci. USA.* 83:4157–4161.
255. Taylor, J. L., J. Jones, S. Sandler, G. Mueller, J. Bedbrook, and P. Dunsmuir. 1987. Optimizing the expression of chimeric genes in plant cells. *Mol. & Gen. Genet.* 210:572–577.
256. Tedder, T. F., M. Streuli, S. F. Schlossman, and H. Saito. 1988. Isolation and structure of a cDNA encoding the B1 (CD20) cell-surface antigen of human B lymphocytes. *Proc. Natl. Acad. Sci. USA.* 85:208–212.
257. Tempel, B. L., Y. Nung Jan, and L. Y. Jan. 1988. Cloning of a probable potassium channel gene from mouse brain. *Nature (Lond.).* 332:837–839.
258. Terwilliger, E., R. Burghoff, R. Sia, J. Sodroski, W. Haseltine, and C. Rosen. 1988. The *art* gene product of human immunodeficiency virus is required for replication. *J. Virol.* 62:655–658.
259. de Thé, H., A. Marchio, P. Tiollais, and A. Dejean. 1987. A novel steroid thyroid hormone receptor-related gene inappropriately expressed in human hepatocellular carcinoma. *Nature (Lond.).* 330:667–670.
260. Thiede, M. A., G. J. Strewler, R. A. Nissenson, M. Rosenblatt, and G. A. Rodan. 1988. Human renal carcinoma expresses two messages encoding a parathyroid hormone-like peptide: evidence for the alternative splicing of a single-copy gene. *Proc. Natl. Acad. Sci. USA.* 85: 4605–4609.
261. Thireos, G., M. Driscoll Penn, and H. Greer. 1984. 5′-Untranslated sequences are required for the translational control of a yeast regulatory gene. *Proc. Natl. Acad. Sci. USA.* 81:5096–5100.
262. Thomas, K. R., and M. R. Capecchi. 1986. Introduction of homologous DNA sequences into mammalian cells induces mutations in the cognate gene. *Nature (Lond.).* 324:34–38.
263. Thomas, K. R., and M. R. Capecchi. 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell.* 51:503–512.
264. Thomas, S. M., R. A. Lamb, and R. G. Paterson. 1988. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell.* 54:891–902.
265. Trono, D., R. Andino, and D. Baltimore. 1988. An RNA sequence of hundreds of nucleotides at the 5′ end of poliovirus RNA is involved in allowing viral protein synthesis. *J. Virol.* 62:2291–2299.
266. Trono, D., J. Pelletier, N. Sonenberg, and D. Baltimore. 1988. Translation in mammalian cells of a gene linked to the poliovirus 5′ noncoding region. *Science (Wash. DC).* 241:445–448.
267. Tyc, K., M. Konarska, H. J. Gross, and W. Filipowicz. 1984. Multiple ribosome binding to the 5′-terminal leader sequence of tobacco mosaic virus RNA. *Eur. J. Biochem.* 140:503–511.
267a. Tzamarias, D., and G. Thireos. 1988. Evidence that the *GCN2* protein kinase regulates reinitiation by yeast ribosomes. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:3547–3551.
268. Van Kranen, H. J., L. van de Zande, C. F. van Kreijl, A. Bisschop, and B. Wieringa. 1987. Cloning and nucleotide sequence of rat ornithine decarboxylase cDNA. *Gene.* 60:145–155.
269. Virbasius, J. V., and R. C. Scarpulla. 1988. Structure and expression of rodent genes encoding the testis-specific cytochrome *c*. *J. Biol. Chem.* 263:6791–6796.
270. Vogel, U. S., R. Dixon, M. Schaber, R. Diehl, M. Marshall, E. M. Scolnick, I. Sigal, and J. B. Gibbs. 1988. Cloning of bovine GAP and its interaction with oncogenic *ras* p21. *Nature (Lond.).* 335:90–93.
271. Volokita, M., and C. R. Somerville. 1987. The primary structure of spinach glycolate oxidase deduced from the DNA sequence of a cDNA clone. *J. Biol. Chem.* 262:15825–15828.
272. Voronova, A. F., H. T. Adler, and B. M. Sefton. 1987. Two *lck* transcripts containing different 5′ untranslated regions are present in T cells. *Mol. Cell. Biol.* 7:4407–4413.
273. Wainwright, B. J., P. Scambler, P. Stanier, E. Watson, G. Bell, C. Wicking, X. Estivill, M. Courtney, A. Boue, P. Pedersen, R. Williamson, and M. Farrall. 1988. Isolation of a human gene with protein sequence similarity to human and murine int-1 and the *Drosophila* segment polarity mutant *wingless*. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1743–1748.
274. Wang, F., L. Petti, D. Braun, S. Seung, and E. Kieff. 1987. A bicistronic Epstein-Barr virus mRNA encodes two nuclear proteins in latently infected, growth-transformed lymphocytes. *J. Virol.* 61:945–954.
275. Wawrousek, E. F., and J. Piatigorsky. 1987. Expression of the δ1-crystallin gene: salt-dependent alteration in the cell-free synthesis of δ-crystallin polypeptides. *Gene.* 58:155–165.
276. Weir, J. P., and B. Moss. 1987. Determination of the transcriptional regulatory region of a vaccinia virus late gene. *J. Virol.* 61:75–80.
277. Werner, M., A. Feller, F. Messenguy, and A. Piérard. 1987. The leader

peptide of yeast gene *CPA1* is essential for the translational repression of its expression. *Cell.* 49:805–813.
278. Weyer, U., and R. D. Possee. 1988. Functional analysis of the p10 gene 5' leader sequence of the *Autographa californica* nuclear polyhedrosis virus. *Nucleic Acids Res.* 16:3635–3653.
279. Whyte, P., K. J. Buchkovich, J. M. Horowitz, S. H. Friend, M. Raybuck, R. A. Weinberg, and E. Harlow. 1988. Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. *Nature (Lond.).* 334:124–129.
280. Wieczorek, D. F., C. W. J. Smith, and B. Nadal-Ginard. 1988. The rat α-tropomyosin gene generates a minimum of six different mRNAs coding for striated, smooth, and nonmuscle isoforms by alternative splicing. *Mol. Cell. Biol.* 8:679–694.
281. Wong, T. C., and A. Hirano. 1987. Structure and function of bicistronic RNA encoding the phosphoprotein and matrix protein of measles virus. *J. Virol.* 61:584–589.
282. Yang, Z., A. J. Korman, J. Cooper, D. Pious, R. S. Accolla, R. C. Mulligan, and J. L. Strominger. 1987. Expression of HLA-DR antigen in human class II mutant B-cell lines by double infection with retrovirus vectors. *Mol. Cell. Biol.* 7:3923–3928.
283. Young, D., K. O'Neill, T. Jessell, and M. Wigler. 1988. Characterization of the rat *mas* oncogene and its high-level expression in the hippocampus and cerebral cortex of rat brain. *Proc. Natl. Acad. Sci. USA.* 85:5339–5342.
284. Zerial, M., D. Huylebroeck, and H. Garoff. 1987. Foreign transmembrane peptides replacing the internal signal sequence of transferrin receptor allow its translocation and membrane binding. *Cell.* 48:147–155.
285. Zhan, X., B. Bates, X. Hu, and M. Goldfarb. 1988. The human FGF-5 oncogene encodes a novel protein related to fibroblast growth factors. *Mol. Cell. Biol.* 8:3487–3495.
286. Zhang, K., W. F. Bosron, and H. J. Edenberg. 1987. Structure of the mouse *Adh-1* gene and identification of a deletion in a long alternating purine-pyrimidine sequence in the first intron of strains expressing low alcohol dehydrogenase activity. *Gene.* 57:27–36.
287. Zitomer, R. S., D. A. Walthall, B. C. Rymond, and C. P. Hollenberg. 1984. *Saccharomyces cerevisiae* ribosomes recognize non-AUG initiation codons. *Mol. Cell. Biol.* 4:1191–1197.